



INSTITUTO POLITÉCNICO  
DE VIANA DO CASTELO

Ana Maria Rodrigues Esteves

# PLATAFORMA WEB PARA MONITORIZAÇÃO DE DADOS EM CONTEXTO ESCOLAR

Mestrado em Engenharia de Software

Trabalho de Projeto efetuado sob a orientação da  
Doutora Sara Maria da Cruz Maia de Oliveira Paiva  
E Co-orientação do  
Doutor Jorge Manuel Ferreira Barbosa Ribeiro

Novembro de 2016



## RESUMO

Tipicamente a tomada de decisões caracteriza-se pela forma rápida, coerente e consistente com que os agentes das organizações assentam as suas decisões estratégicas para melhorar o desempenho dos serviços operacionais e aumentar a rentabilidade das suas organizações. Por sua vez, as Escolas/Agrupamentos possuem uma imensidão de informação que não sendo devidamente catalogada e tratada, podem induzir os órgãos de gestão a tomadas de decisão menos corretas.

Este projeto surgiu da crescente atenção e relevância que é dada aos dados de uma instituição, sendo imperativo que as tomadas de decisão sejam corretas e céleres mediante uma excessiva quantidade de informação existente e da necessidade de analisar essa informação através de várias perspetivas de análise (conjugando vários fatores de informação), permitindo aos agentes de decisão terem acesso à informação através de várias perspetivas de análise. No estudo de diagnóstico realizado no âmbito deste trabalho foram observadas fortes limitações ao nível da compilação da informação, tendo sido identificada a necessidade de desenvolver uma plataforma integrada de gestão de informação, capaz de agregar grandes volumes de dados. Neste sentido, pretendeu-se projetar e desenvolver uma aplicação de tratamento estatístico de dados de alunos, professores e funcionários de um Agrupamento de Escolas, em que o sistema deveria fornecer ao utilizador uma visão global dos dados numa estrutura bem planificada.

Neste contexto, este projeto foi desenvolvido no **Agrupamento de Escolas Pintor José de Brito** e para colmatar a necessidade de agregar dados e permitir a sua exploração em várias dimensões, o sistema proposto é baseado nos Sistemas de Apoio à Decisão, em particular nas plataformas de *Business Intelligence* (BI). Tais plataformas permitem agregar dados de várias fontes permitindo a exploração analítica sobre várias dimensões auxiliando os agentes de decisão a complementarem a informação das suas fontes de dados. Este trabalho seguiu a metodologia de investigação *Design Research*, cujo objetivo é orientar e validar a construção de artefactos, materializando-se em termos práticos no desenvolvimento e implementação de uma plataforma de BI, sendo elaborada uma proposta de um sistema de apoio à decisão. A nível técnico foi seguida a metodologia de *Kimball* e foi utilizado o *Pentaho Community Edition* como ferramenta *Open Source* de BI, que permitiu conferir várias características ao projeto, como a integração e dados e a versatilidade na visualização da informação segundo várias perspetivas de análise. Posteriormente na parte final do trabalho e baseando-se na metodologia *Design Research* procedeu-se à exploração, validação e divulgação do sistema.

Palavras-Chave: Business Intelligence, Data Warehouse, Data Mart, Sistemas de Suporte à Decisão, Escolas, Open Source, Pentaho

## ABSTRACT

Typically decision making is characterized by the fast, coherent and consistent way that organizations' agents base their strategic decisions to improve the performance of operational services and increase the profitability of their organizations. In turn, Schools/School Groupings have an immense quantity of information which, if not properly catalogued and treated, may induce management bodies to incorrect decisions. Schools should not neglect the volume of information they have and attend to the need of an effective information system to treat and analyse their data. Information has to flow from the school's management body to the staff, teachers and students, and vice-versa. This project emerged of the growing attention and relevance given to institutional data. It is crucial that decision-making is correct and speedy in the face of the current excessive quantity of information and the need to analyse it in different perspectives (combining several factors of information). The diagnostic study made in this project revealed strong limitations in compiling information, showing the need to develop an integrated platform of information management. Therefore, the goal was to project and develop an application to statistically treat the data of students, teachers and staff of a school grouping, where the system should provide the user with a global vision of the data in a well planned structure. The main characteristics that the application should possess are usability and reliability of the information provided to the user.

In this context, the project was developed in the **School Pintor José de Brito**. To solve the need to aggregate data and allow their exploration in several dimensions, the system proposed is based in Decision Support Systems, particularly Business Intelligence platforms which allow to aggregate data from several sources and their analytical exploration in several dimensions thus helping decision agents to complement the information in their data sources.

This work was based in the methodology of Design Research, whose goal is to guide and validate the creation of artefacts, materializing them in the development and implementation of a business intelligence platform and the proposition of a decision support system. Technically, Kimball methodology was used, and the Open Source Business Intelligence tool was Pentaho Community Edition which conferred it several characteristics. At the final stage of the work, based in Design Research methodology, the exploration, validation and publicizing of the system were performed.

Key-Words: Business Intelligence, Data Warehouse, Data Mart, Decision Support Systems, School, Open Source, Pentaho

## AGRADECIMENTOS

À Professora Sara Paiva e ao Professor Jorge Ribeiro pelo constante apoio e por acreditar na realização do presente trabalho.

Aos meus familiares, pela ajuda e incentivo.

Ao Agrupamento de Escolas Pintor José de Brito pela oportunidade de realizar este trabalho e pelas sugestões e apoio demonstrado nas reuniões de trabalho.

Aos meus colegas de mestrado pela paciência, parceria e incentivo, que foram determinantes para a conclusão deste projeto.

A todos os outros, muitos, a quem devo o suficiente para ter conseguido efetuar a presente dissertação.



# ÍNDICE

1. Introdução.....	7
1.1 Contexto e Motivação .....	7
1.2 Metodologia de Investigação e Objetivos.....	8
1.3 Estrutura do Documento.....	10
2. Sistemas de Suporte à decisão .....	11
2.1 Introdução .....	11
2.2 Business Intelligence .....	13
2.3 Ferramentas/Plataformas de Business Intelligence.....	28
2.4 Metodologias de Implementação de Plataformas de BI.....	32
2.5 Considerações Finais .....	34
3. Caso de Estudo - Análise de Requisitos .....	35
3.1 Introdução .....	35
3.2 A Instituição de Ensino Alvo do Estudo.....	35
3.3 Trabalhos Relacionados .....	39
3.4 Requisitos .....	42
3.5 Modelo de Domínio.....	45
3.6 Casos de Uso .....	46
3.7 Especificação da Prototipagem .....	47
3.8 Considerações Finais .....	48
4. Plataforma de BI-Desenvolvimento .....	49
4.1 Introdução .....	49
4.2 Projeto Técnico da Arquitetura .....	49
4.3 Seleção e Escolha das Ferramentas .....	50
4.4 Desenvolvimento do Sistema ETL e da Estrutura da Plataforma de BI .....	52
4.5 Visualização de Dados .....	70
4.6 Manutenção e Crescimento .....	74
4.7 Apreciação e Avaliação Global da Plataforma .....	74
5. Avaliação dos Resultados do Trabalho .....	76
5.1 Introdução .....	76

5.2 Testes .....	76
5.3 Discussão dos Resultados.....	80
5.4 Dificuldades Encontradas na Implementação do Sistema BI.....	81
5.5 Notas Finais .....	81
6. Conclusões e Trabalho Futuro .....	82
6.1 Conclusões.....	82
6.2 Trabalhos Futuros.....	83
Referências.....	84
Anexos.....	87
Anexo 1 – Manual de Instalação e Configuração do Sistema .....	87
Anexo 2 – Inquéritos de Usabilidade do Sistema.....	90



# ÍNDICE DE FIGURAS

FIGURA 1: ARQUITETURA GENÉRICA DE UM SISTEMA DE BI .....	15
FIGURA 2: AMBIENTE DW.....	16
FIGURA 3: PROCESSO ETL.....	17
FIGURA 4: EXEMPLO DA TRANSFORMAÇÃO DIM_FUNCIONÁRIOS .....	17
FIGURA 5: VISÃO DW E DM .....	20
FIGURA 6: COMPONENTES DO MODELO DIMENSIONAL .....	21
FIGURA 7: MODELO EM ESTRELA .....	22
FIGURA 8: MODELO EM FLOCO DE NEVE .....	22
FIGURA 9: CUBO GENÉRICO .....	25
FIGURA 10: EXEMPLO <i>DASHBOARD</i> .....	28
FIGURA 11: PENTAHO BI <i>SUITE</i> .....	30
FIGURA 12: PLATAFORMA PENTAHO OPEN BI <i>SUITE</i> .....	31
FIGURA 13: CICLO DE VIDA DA METODOLOGIA DE KIMBALL .....	32
FIGURA 14: DISTRIBUIÇÃO DE DOCENTES POR NÍVEL DE ENSINO .....	36
FIGURA 15: DISTRIBUIÇÃO DE DOCENTES POR VÍNCULO.....	37
FIGURA 16: DISTRIBUIÇÃO DE FUNCIONÁRIOS POR CATEGORIA.....	37
FIGURA 17: DISTRIBUIÇÃO DAS TURMAS DO AGRUPAMENTO POR ANO.....	37
FIGURA 18: DISTRIBUIÇÃO DO Nº DE ALUNOS POR CICLO DE ENSINO.....	38
FIGURA 19: <i>DASHBOARD</i> EDUCACIONAL - INFOESCOLAS .....	42
FIGURA 20: MODELO DOMÍNIO .....	45
FIGURA 21: DIAGRAMA USE CASE (ADMIN) .....	46
FIGURA 22: DIAGRAMA USE CASE (USER) .....	47
FIGURA 23: PROTÓTIPO DA APLICAÇÃO INFOPINTOR .....	48
FIGURA 24: ARQUITETURA DO SISTEMA SEGUNDO KIMBALL.....	50
FIGURA 25: TECNOLOGIAS USADAS NO DESENVOLVIMENTO DA APLICAÇÃO INFOPINTOR .....	51
FIGURA 26: ESQUEMA DO PROCESSO DE DESENVOLVIMENTO DA APLICAÇÃO .....	52
FIGURA 27: DM CICLO ENSINO .....	56
FIGURA 28: DM FUNCIONÁRIO.....	56
FIGURA 29: DM INSCRITOS.....	57
FIGURA 30: ESQUEMA DE TRANSFORMAÇÃO ALUNOS .....	58
FIGURA 31: ESQUEMA DE TRANSFORMAÇÃO FUNCIONÁRIOS.....	58
FIGURA 32: CRIAR DIMENSÃO ALUNO (DIM_ALUNO) .....	59
FIGURA 33: CRIAR DIMENSÃO ANO ESCOLAR (DIM_ANO_ESCOLAR) .....	59
FIGURA 34: CRIAR DIMENSÃO CATEGORIA .....	60
FIGURA 35: CRIAR DIMENSÃO TEMPO SERVIÇO .....	60
FIGURA 36: CRIAR DIMENSÃO FAIXA ETÁRIA .....	60
FIGURA 37: CRIAR FACTUAL CICLO ENSINO .....	61
FIGURA 38: CRIAR FACTUAL INSCRITOS .....	61
FIGURA 39: CRIAR FACTUAL FUNCIONÁRIO .....	62
FIGURA 40: JOB DIMENSÕES .....	62
FIGURA 41: JOB FACTUAIS.....	63
FIGURA 42: TRANSFORMAÇÃO REFRESH MONDRIAN .....	63
FIGURA 43: <i>JOG</i> GERAL.....	63
FIGURA 44: SCRIPT GENÉRICO .....	64
FIGURA 45: ESTRUTURA CONSTRUÇÃO DE UM CUBO OLAP .....	65
FIGURA 46: INSTALAÇÃO DO SAIKU .....	67

FIGURA 47: EXPLORAÇÃO NO SAIKU .....	68
FIGURA 48: AMBIENTE DE TRABALHO SAIKU – MATRIZ CRUZADA .....	68
FIGURA 49: AMBIENTE DE TRABALHO SAIKU – EXPLORAÇÃO GRÁFICA .....	69
FIGURA 50: AMBIENTE DE TRABALHO SAIKU – ESTATÍSTICAS BÁSICAS .....	69
FIGURA 51: EXEMPLO <i>QUERY</i> MDX E RESPECTIVO RESULTADO .....	70
FIGURA 52: COMPONENTES DA <i>SUITE</i> CTOOLS.....	71
FIGURA 53: VISTA DO MENU <i>LAYOUT</i> .....	72
FIGURA 54: VISTA DO MENU <i>COMPONENTS</i> .....	72
FIGURA 55: VISTA DO MENU <i>DATA SOURCES</i> .....	73
FIGURA 56: EXEMPLO DE <i>QUERY</i> MDX.....	73
FIGURA 57: <i>LAYOUT</i> FINAL DA APLICAÇÃO INFOPINTOR .....	74
FIGURA 58: GESTÃO DE UTILIZADORES .....	77
FIGURA 59: GESTÃO DE PERFIS DE UTILIZADOR.....	77
FIGURA 60: TESTES ETL.....	78
FIGURA 61: VALIDAÇÃO GRÁFICA DOS CUBOS NO SAIKU .....	79
FIGURA 62: VALIDAÇÃO ATRAVÉS DE <i>QUERY</i> SQL.....	79
FIGURA 63: ANÁLISE DO NÚMERO FUNCIONÁRIOS .....	80
FIGURA 64: VARIÁVEIS DE SISTEMA .....	87
FIGURA 65: DOWNLOAD PENTAHO.....	87
FIGURA 66: CONSOLA DURANTE EXECUÇÃO PENTAHO .....	88
FIGURA 67: JANELA DE LOGIN .....	88
FIGURA 68: MENU FILE .....	88
FIGURA 69: LIGAÇÃO À BD.....	89

## ÍNDICE DE TABELAS

TABELA 1: COMPARAÇÃO ENTRE FERRAMENTAS DE BI.....	29
TABELA 2: TABELAS DE DIMENSÃO DO DW.....	55
TABELA 3: TABELAS DE FACTOS DO DW .....	55
TABELA 4: TABELA DE MÉTRICAS.....	55
TABELA 5: VALIDAÇÃO DOS CUBOS NO SAIKU .....	79
TABELA 6: RESULTADO DA VALIDAÇÃO <i>QUERY</i> SQL.....	79

## LISTA ACRÓNIMOS

QZP	Professor do Quadro Zona Pedagógica
RF	Requisito Funcional
RNF	Requisito Não Funcional
BI	Inteligência de Negócios (do inglês Business Intelligence)
DW	Armazém de Dados (do inglês Data Warehouse)
ETL	Extração, Transformação e Carga (do inglês Extraction, Transformation and Loading)
DM	Data Mart
OLAP	Processamento Analítico em Tempo Real (do inglês Online Analytical Processing)
PDI	Open Source – Pentaho Data Integration
SGBD	Sistema de Gestão de Bases de Dados
SQL	Structured Query Language
BD	Base de Dados
XML	eXtensible Markup Language
GPV	Gestão de Pessoal e Vencimentos
ERP	Enterprise Resources Planning
CDA	Community Data Access
CDE	Community Dashboard Editor
CDF	Community Dashboard Framework
MDX	Multidimensional Expressions
OS	Open Source
ENES	Exames Nacionais do Ensino Secundário
ENEB	Exames Nacionais do Ensino Básico
PFEB	Provas Finais do 1.º ciclo



# 1. INTRODUÇÃO

Este documento apresenta o trabalho desenvolvido ao longo do projeto de dissertação do Mestrado em Engenharia de Software. Este capítulo encontra-se dividido em 3 secções. Na primeira (1.1) é apresentada uma breve contextualização e motivação do projeto. Na segunda secção (1.2) é apresentada a metodologia adotada e os objetivos do projeto. Conclui-se o capítulo com a apresentação da estrutura deste documento.

## 1.1 CONTEXTO E MOTIVAÇÃO

Nos dias de hoje, verifica-se diariamente, uma crescente quantidade de informação, tornando-se necessário uma gestão imperativa e eficiente do conhecimento por parte das organizações com vista à diminuição dos custos e o aumento da eficiência interna e externa nas suas tarefas de gestão. Por sua vez um Sistema de Apoio à Decisão é tipicamente caracterizado por um ou mais sistemas informáticos que suportam as atividades de tomada de decisão nas empresas ou organizações, devendo os mesmos sistemas facilitarem o acesso rápido e integrado das diferentes fontes de informação que influenciam a tomada de decisões.

Tendo em conta os aspetos referidos anteriormente, no século XXI, houve uma explosão na quantidade de dados gerados em todas as áreas do conhecimento, fazendo com que a tarefa mais difícil seja a sua análise e exploração. O tratamento de informações em grandes bases de dados requer um procedimento que obriga à utilização de técnicas e ferramentas que facilitem as ações de recolha, análise e utilização das informações, de forma ágil e confiável (Hall, 2011); (Koomey, 2011). Com o crescente volume de dados gerados pelos atuais sistemas de informação, é cada vez mais evidente a dificuldade da sua manipulação por forma a extrair informação útil. Uma correta e precisa interpretação da informação conduz a uma tomada de decisão mais rápida, que, nos dias de hoje, é fundamental para o sucesso de uma organização.

No contexto da gestão escolar, os Diretores dos Agrupamentos têm a responsabilidade de compartilhar uma quantidade de dados publicados e relacionados, para além de transmitir as informações de interesse para toda a comunidade educativa. A estruturação

e disponibilização das informações geradas nos seus diversos sistemas, de modo a que os Diretores e a sua equipa as utilizem como recurso estratégico e possam obter vantagem competitiva, é um dos desafios deste trabalho. Essas informações, exploradas de forma eficiente, podem ser cruciais para melhorar os processos de ensino, sendo que um dos problemas na utilização de dados educacionais para melhorar tais processos vincula-se à aquisição de informação.

No **Agrupamento de Escolas Pintor José de Brito**, adiante designado por APJB comprovou-se que não é possível compilar, em tempo, útil, as informações disponíveis nas várias plataformas usadas, sendo que o Agrupamento não possui uma plataforma organizada de informação estatística, existindo muitas vezes a necessidade de navegar pela informação, de modo a comparar diversos fatores cruciais para dar resposta ao Projeto Educativo. Neste sentido, o trabalho alvo passou por condensar uma série de indicadores importantes para o tema em questão (desenvolvimento de uma plataforma web para a monitorização de dados em contexto escolar) e procurar traduzi-los de forma gráfica, implementando uma interface denominada “InfoPintor”. Esta abordagem caracteriza-se pela forte componente visual, dinamismo e possibilidade de manipulação por parte do utilizador.

Por sua vez, de dia para dia, são analisadas maiores quantidades de dados, com mais variáveis a integrarem o sistema. Desta forma, torna-se necessária a utilização de mecanismos que permitam analisar e tratar dados, de forma a ser possível tomar uma decisão racional, baseada em dados tratados e consistentes, e continuar a garantir o sucesso do meio em que esta informação é utilizada. Em suma, as vantagens destas aplicações são evidentes, tornando a tomada de decisões por parte das chefias mais fácil e produtiva.

## **1.2 METODOLOGIA DE INVESTIGAÇÃO E OBJETIVOS**

Neste trabalho pretende-se desenvolver um sistema de BI para um Agrupamento de Escolas recorrendo a tecnologias *open source*. Avaliar as possibilidades de criar e manter um sistema BI, de baixo custo, fiável e capaz de satisfazer as necessidades de uma organização pública local é o objetivo deste trabalho. Neste contexto, este trabalho não se centrou tipicamente num trabalho puro de investigação científico, mas antes em

desenvolver um projeto aplicado a um contexto específico, no sentido de por um lado, melhorar o caso de estudo e por outro, aplicar os vários conhecimentos adquiridos nas várias unidades curriculares do curso de Mestrado.

Para o desenvolvimento deste projeto, seguiu-se a orientação da metodologia de investigação *Design Research* (Manson, 2006) (Järvinen, 1997) a qual pressupõe a ação do investigador numa determinada realidade, permitindo-lhe compreender um problema e mediante isso propor a construção de uma solução (artefacto) e proceder ao teste da mesma, passando o investigador de um mero observador para um indivíduo que age no contexto investigado/aplicado, procurando compreender uma determinada realidade, utilizando o seu potencial criativo para criar soluções para problemas ou necessidades reais (Wang & Hanna, 1997).

Tradicionalmente a metodologia *Design Research* encontra-se dividida em 5 passos: Consciencialização (a qual corresponde à compreensão da problemática envolvida, sendo os principais resultados a definição e formalização do problema a ser solucionado), Sugestão (centrada na definição dos objetivos para a solução e nas atividades de desenvolvimento de uma solução), Desenvolvimento (processo de construção do artefacto pelo pesquisador), Avaliação (o artefacto deve ser analisado e testado de acordo com as condições estabelecidas para validação; esta etapa contribui ao processo de melhoria do artefacto construído) e a Conclusão (centra-se na avaliação da solução para a resolução/minimização do problema). Para a avaliação final do artefacto é necessário que, em cada etapa do método *Design Research*, sejam realizadas avaliações parciais dos resultados, sendo importante confirmar que a pesquisa está sendo feita no sentido dos objetivos propostos.

Neste contexto, na etapa de consciencialização, consideramos o facto de no agrupamento APJB não ser possível compilar, em tempo útil, as informações disponíveis nas várias plataformas e não ter uma plataforma organizada de informação estatística que auxilie os agentes de decisão a complementarem as suas decisões com informação relevante/importante que até então não a usavam. Neste sentido e seguindo a metodologia, a sugestão foi a de criar uma plataforma web unificada para a monitorização de dados. A fase de desenvolvimento e as restantes serão detalhadas mais adiante neste documento.

### **1.3 ESTRUTURA DO DOCUMENTO**

O presente documento está estruturado em seis capítulos. No primeiro capítulo apresenta-se o enquadramento deste trabalho ao nível da sua contextualização, âmbito em que se insere, escola a que se destina, metodologia de investigação bem como os objetivos traçados para a sua realização. No segundo capítulo apresenta-se a contextualização dos Sistemas de Suporte à Decisão. No terceiro e quarto capítulo apresenta-se o caso de estudo em termos de especificação e de requisitos, assim como a implementação do mesmo. No quinto é feita uma avaliação do trabalho apresentado ao Agrupamento, são apresentados os resultados de alguns testes e validação de dados. No capítulo 6 são apresentadas as conclusões do trabalho realizado bem como as propostas para futuros trabalhos.

Por fim apresentam-se as referências bibliográficas que sustentaram a parte teórica do projeto e em anexo o manual de instalação do sistema bem como o inquérito aplicado aos órgãos de decisão do Agrupamento com os respetivos resultados apurados.



## 2. SISTEMAS DE SUPORTE À DECISÃO

Neste capítulo é apresentada de uma forma geral a revisão da literatura e o estado da arte dos termos e metodologias que foram utilizados ao longo do desenvolvimento desta dissertação de mestrado. Pretendeu-se que as metodologias estudadas fossem o suporte científico dos métodos e procedimentos utilizados ao longo da realização do projeto.

Este capítulo encontra-se dividido em 5 secções. Na primeira secção (2.1) é abordada a introdução aos sistemas de apoio à decisão na educação. Na segunda (2.2) é explorado o BI e as diferentes fases inerentes a este. Na terceira (2.3) são abordadas as ferramentas de *BI Open Source* (OS). Na quarta (2.4) é abordada a metodologia de implementação da plataforma de BI para o caso de estudo. Na quinta e última secção são apontadas as considerações finais do capítulo.

### 2.1 INTRODUÇÃO

As organizações, sejam elas públicas ou privadas, estão sob constantes pressões e, como consequência, têm de responder rapidamente às condições de mudança, ser inovadoras no atendimento às necessidades dos seus clientes e no modo de atuarem perante o mercado. Estes fatores tornam-se críticos para o sucesso. Isto implica que as organizações sejam ágeis e céleres a tomar decisões, sejam elas estratégicas, táticas ou operacionais (Vercellis, 2009). No entanto, para tomar essas decisões é necessário considerar uma grande quantidade de dados, de informação e de conhecimento. A informação quando transformada em conhecimento, gera uma mais-valia fundamental para a tomada de decisões. Da mesma forma, a escassez de informação pode conduzir a erros e à perda de oportunidades. Com um conjunto de dados históricos e centralizados, é mais facilmente possível traçar metas, tornando o processo de tomada de decisão mais seguro e fiável.

O APJB apenas dispõe da página oficial e da página do *Facebook*, no entanto não responde a nenhum tipo de dados estatísticos. De nada adianta ter uma base de dados com milhares de informações se não forem devidamente tratadas e visualizadas. Existe necessidade de demonstrar resultados, metas, comparativos e históricos. O principal

objetivo deste projeto é desenvolver e projetar uma aplicação web centrada em dados estatísticos do Agrupamento relativamente aos alunos, professores e funcionários. Por sua vez, constata-se que a gestão de um Agrupamento em que as tomadas de decisão dependem da perceção eficaz de uma grande quantidade de fatores pode ser suportada por interfaces mais ágeis e com mais informação que os apoiem em decisões com informação mais consistente segundo várias vertentes de análise.

Neste capítulo apresentam-se os elementos teóricos essenciais que envolvem um sistema baseado em *Data Warehouse* (DW), bem como as definições das técnicas e ferramentas que foram utilizadas. Segundo (Chowdhury, 2014) o sucesso depende da forma como se organiza e proporciona o acesso à informação necessária às pessoas, organizações, empresas e sociedade em geral para ajudar na tomada de decisões. Ao longo dos anos, a informação tornou-se um dos ativos mais importantes e valiosos das organizações. Na visão de (Davenport, 2002), num ambiente organizacional e institucional, a informação representa um elemento de poder; no campo do processo da decisão, quanto mais confidencial for uma informação, maior é o seu valor. Já para (Abreu, 2013) a informação tem um valor significativo e constitui um poder para quem a possui, seja pessoa, seja instituição. Por outro lado, a atividade de apoio à decisão não se limita à aplicação de um método multicritério com vista à resolução de uma situação decisória determinada (...) (Costa, 1986), podendo existir um conjunto alargado de possíveis soluções, no entanto, após a decisão do caminho a percorrer, este deve ser levado a cabo superando as dificuldades e minimizando o seu impacto.

Numa outra perspetiva de análise de informação, tipicamente os agentes de decisão utilizam folhas de cálculo para reunirem a informação que necessitam, mas na maioria das vezes sem tratamento ou sem agregar, integrar, ou interligar com outra de forma direta a informação que dispõe. Por sua vez, os denominados “*Dashboards*” ou “quadros” ou “mapas resumo” de informação de apoio à decisão vieram revelar-se uma ferramenta auxiliar de enorme vantagem na tomada à decisão. As plataformas que disponibilizam os *Dashboards*, tipicamente representadas num Website exigem estudos prévios de boa usabilidade. Se avaliarmos os componentes da usabilidade (aprendizagem, eficiência, lembrança, erros, satisfação), é notório que o utilizador é sempre o centro das atenções, não podendo ser esquecido neste processo criativo. O *Dashboard* necessita igualmente de *design* para poder alavancar todo o seu potencial. Seria muito difícil, se não quase

impossível, conseguir num único ecrã transmitir toda a informação relevante do Agrupamento.

## **2.2 BUSINESS INTELLIGENCE**

Hoje em dia, a informação e o conhecimento representam a riqueza fundamental de uma organização. As empresas tentam utilizar essa riqueza para ganhar vantagens competitivas na tomada de decisões importantes (Ferreira, Santos, & Serra, 2010). Para tal existe a necessidade de se obter informações detalhadas de modo a que as organizações possam analisar essa mesma informação, e assim, realizarem uma gestão mais eficiente da mesma. Neste contexto, faz sentido referir aos sistemas de BI, um conceito caracterizado por um sistema integrado ou conjunto de conceitos e métodos para melhorar a tomada de decisão baseada na recolha, organização, análise e apresentação dos dados, de tal maneira a apoiar uma melhor decisão. No fundo o BI é um “método”, “abordagem”, “conjunto de ferramentas” que permitem agrupar as informações de modo a que seja possível tomar uma decisão mais fundamentada. Trata-se de um processo que junta a análise de dados e ferramentas de análise de decisão para disponibilizar as informações certas para as pessoas certas em toda a organização, com o objetivo de melhorar a tomada de decisões estratégicas. Em geral, o BI transforma a informação em conhecimento e tem a capacidade de disponibilizar informação correta ao utilizador em qualquer altura e em qualquer lugar (Damiani, 2009).

Fue Howard Dresner, o primeiro que intitulou, em 1989, o termo BI, definiu-o como:

“BI é um processo interativo para explorar e analisar informação estruturada sobre uma área (normalmente armazenada num DW), para descobrir tendências e padrões, a partir dos quais derivam ideias e se extraem conclusões. O processo de BI inclui a comunicação do que foi apurado e o processo de efetuar as devidas alterações. As áreas incluem clientes, fornecedores, produtos, serviços e concorrentes.”

Segundo (Xu, Zeng, Shi, He, & Wang, 2007), BI denomina um conjunto de ferramentas e soluções tecnológicas desenhadas para que os utilizadores possam extrair de maneira eficiente informação útil para apoiar o processo de tomada de decisão, obtendo assim vantagens competitivas sobre outras organizações.

Neste sentido e tendo em conta o sucesso da ampla aplicabilidade da utilização de ferramentas de BI no apoio à decisão e da grande abrangência em várias áreas de atividade, considera-se uma poderosa ferramenta para incrementar valor às organizações.

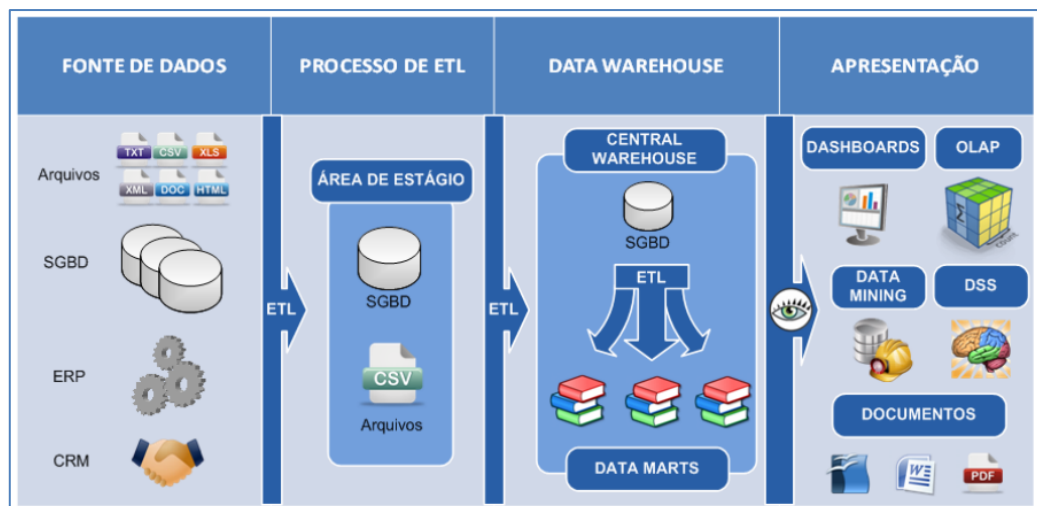
### **2.2.1 NECESSIDADE DO BI E ESTRUTURA ARQUITETURAL**

Como foi referido anteriormente, existe a necessidade de tomada de decisões rápidas, focada na agilidade e análise dos dados, para que as organizações sejam competitivas. Contudo, as restrições de tempo e a desorganização dos dados limitam a capacidade dos gestores em tomar decisões eficientes que origina a que, muito frequentemente, a diferença entre a decisão certa e a errada é a informação com que os agentes de decisão se baseiam.

Perante cenários de descentralização e heterogeneidade dos dados, o grande desafio de hoje centra-se na sua integração, interpretação e a sua transformação em informação relevante para o sistema, possibilitando, com a devida análise, a criação de conhecimento. Os sistemas de informação tradicionais (SGBD, ERP, entre outros), apresentam, na sua maioria, estruturas complexas onde se torna complicado extrair informação e por vezes impossível extrapolar o conhecimento obtido para o dia-a-dia das organizações. Em termos gerais, as principais características que limitam estes sistemas caracterizam-se:

- Pela rigidez na extração e visualização dos dados;
- Pelo tempo de resposta demorado;
- Pelas lacunas nos dados (ex.: inconsistências, dados omissos, não relacionados, entre outros).

Nos dias de hoje não interessa o tamanho da instituição, mas perante as necessidades do mercado, existe a necessidade do BI auxiliar nas mais diferentes tarefas, permitindo perspetivar estratégias robustas atendendo às várias dimensões de análise de informação. Os componentes que constituem um sistema de BI listam-se abaixo e representam-se na figura 1.



**Figura 1: Arquitetura Genérica de um Sistema de BI**

Fonte: Adaptado de Bouman e Dongen (2010)

A figura anterior ilustra uma arquitetura genérica de um sistema de BI, o qual é caracterizado por:

- Fontes de informação diversas que vão alimentar o sistema;
- Processo ETL que extrai, transforma e carrega os dados das fontes de informação até ao DW;
- DW que reúne os dados da organização;
- Motor *On-line Analytical Processing* (OLAP) que permite capacidade de cálculo, consultas e análise a grandes volumes de dados;
- Ferramentas *Front-end*, são as ferramentas de visualização propriamente ditas. Estas permitem que o utilizador final proceda à navegação pelos dados para uma análise consistente.

### **2.2.2 O PROCESSO EXTRAÇÃO, TRANSFORMAÇÃO E CARREGAMENTO DE DADOS (ETL)**

No contexto de BI é frequente referir-se ao processo de ETL, como sendo responsável pela extração, transformação e limpeza dos dados provenientes dos diversos sistemas organizacionais (sistemas *Online Transaction Processing*) e pelo seu carregamento, geralmente, para um DW ou Data Mart (DM). O processo de ETL é considerado tipicamente o mais lento e crítico no processo de construção de um DW dada a grande quantidade de dados a serem processados. Além disso, existe uma complexa combinação

de processos e tecnologias que consomem uma parcela significativa do tempo inerente a estas etapas (El-Sappagh, Hendawi, & Bastawissy, 2011). Na figura seguinte ilustra-se o enquadramento do processo ETL num ambiente de Sistemas de *Data Warehousing* e geração de estruturas de apoio à agregação de dados.

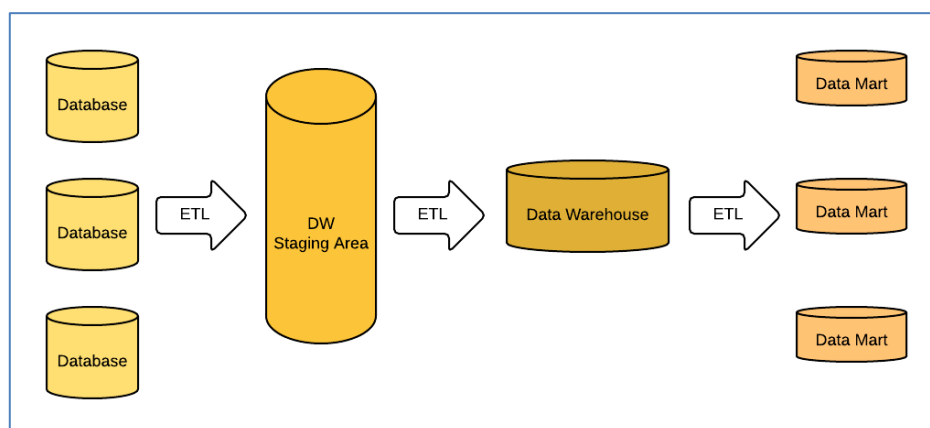


Figura 2: Ambiente DW

Descrevendo com mais detalhe as etapas do processo ETL, segundo (Vassiliadis, 2009), este processo envolve:

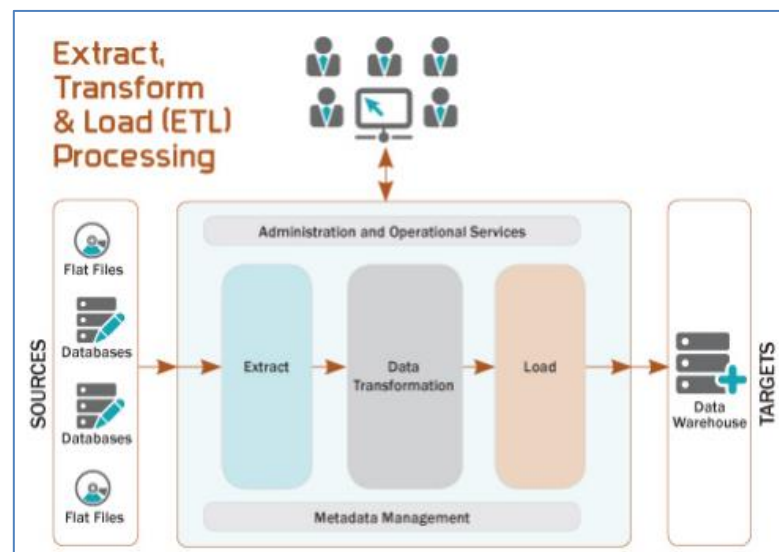
- a extração dos dados apropriados de fontes externas diversas;
- o transporte dos dados para uma área de preparação - *Data Staging Area*<sup>1</sup>;
- a transformação dos dados para atender às necessidades de negócios;
- separar e limpar os registos problemáticos para garantir que as regras do negócio e as restrições da BD são respeitadas e,
- o carregamento dos dados limpos e transformados no DW.

O ETL é importante pois corresponde à forma pela qual os dados são efetivamente carregados no DW, sendo selecionados, pré-processados, tratados, mapeados em termos de equivalência de informação e integrados nos repositórios de suporte. Com as mudanças no negócio, as necessidades do DW também mudam, por isso os processos ETL devem ser desenhados para que sejam fáceis de modificar. Assim, um processo ETL sólido,

---

<sup>1</sup> “Representa um armazenamento intermédio dos dados, facilitando a integração dos dados de origem antes da sua atualização no DW” (Machado, 2004). Esta área armazena os dados em bruto extraídos dos diferentes sistemas. É ainda nesta camada que os dados sofrem transformações: (conversões, junção de dados provenientes de diferentes fontes, criação de chaves substitutas, entre outros.) (Thorlund, 2010).

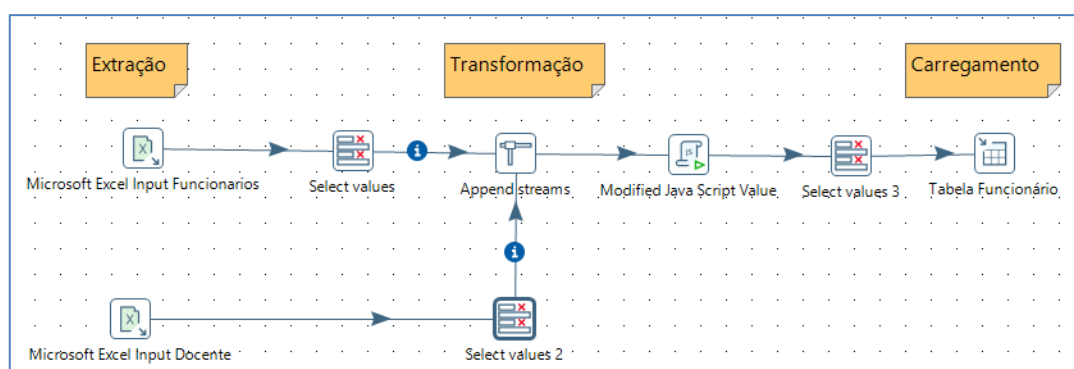
bem desenhado e documentado é necessário para o êxito de um projeto de DW (El-Sappagh, Hendawi, & Bastawissy, 2011).



**Figura 3: Processo ETL**

Fonte: Rick Sherman (2009)

No que se refere à extração dos dados de fontes internas e externas, durante este subprocesso são extraídos os dados necessários, e somente os necessários, a partir das fontes de informação. Para tal, existe a necessidade de averiguar, de entre as fontes disponíveis, quais as melhores e quais as que mais se adequam ao sistema que está a ser planeado. Esta etapa é de extrema importância podendo ser executada com recurso a ferramentas de ETL ou manualmente recorrendo a linguagens de programação para o efeito.



**Figura 4: Exemplo da Transformação Dim\_Funcionários**

Em termos de limpeza dos dados extraídos, este subprocesso é um passo intermédio entre o subprocesso da extração e o subprocesso da transformação. O processo de limpeza e transformação dos dados que serão carregados num sistema de DW serve para corrigir imperfeições ou anomalias das BD, sendo o mais importante nesta fase é a coerência e compatibilidade dos dados. Como exemplo de um processo de transformação, sendo que os dados podem ser provenientes de sistemas diversos, é necessário nesta fase padronizar os diferentes formatos.

Por exemplo, alguns sistemas de informação armazenam o campo “estado civil” no formato: "C" para Casado e "S" para Solteiro. No entanto, outros sistemas podem armazenar no formato "1" para Casado e "2" para Solteiro e assim sucessivamente. Aplicando um processo ETL, os dados são transformados e mantidos num formato uniforme.

No que se refere à carga de dados no DW, este processo está encarregue de carregar os dados resultantes dos processos anteriores no DW. Assegura também a estruturação física dos dados num conjunto de modelos denominados modelos dimensionais (esquemas em estrela).

Em termos de atualizações, as organizações estão em constante mudança e a qualquer momento são realizadas alterações nas suas BD. Estas modificações devem ser atualizadas para o DW a fim de estarem em concordância com as fontes de origem dos dados. A maioria das organizações possui sistemas programados para detetarem automaticamente alterações sofridas nas fontes de dados.

### **2.2.3 DATA WAREHOUSE**

Neste tipo de ambientes de suporte à decisão, os dados resultantes do processo de ETL normalmente são carregados em estruturas multidimensionais denominadas de DW. Estas estruturas são orientadas por temas, onde os dados relativos a diferentes assuntos são armazenados numa ou mais tabelas (denominadas de DM). Uma das atividades mais significativas no âmbito do BI consiste no desenho e construção dos repositórios de dados - o DW.

(Inmon, 1996) define objetivamente o DW e salienta as suas características: integrado (os dados são padronizados e uniformizados); variável no tempo (cada entrada



corresponde a um determinado momento específico no tempo); e não volátil (os dados armazenados não são removidos). Por outras palavras, após a procura de informações úteis em grandes bases de dados, provenientes de diferentes aplicações, quer internas ou externas, elas são estruturadas e armazenadas num repositório central, o DW (March & Hevner, 2007), podendo ser armazenados numa etapa intermédia, em DM, que consolidam pequenas partes de informações relacionadas entre si, isto é, em cada DM há informação relativa a uma área específica.

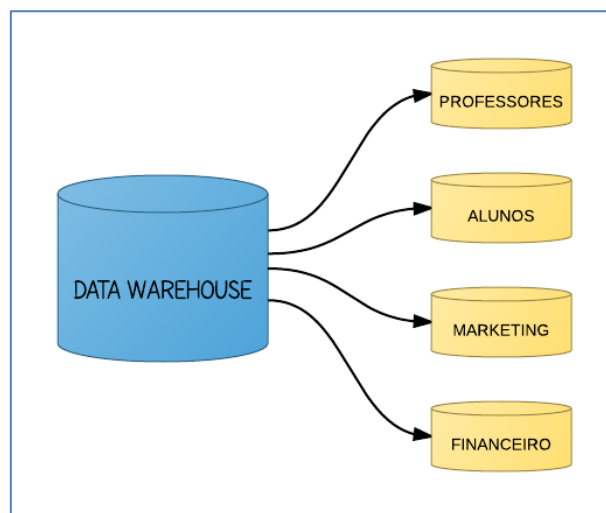
Existem várias abordagens de metodologias a seguir para a construção de um DW e com aplicabilidades diferentes, como Ralph Kimball e Bill Inmon, que divergem quanto à melhor estratégia de implementação. Inmon é defensor da abordagem *top-down* e Kimball da *bottom-up*. Na arquitetura *bottom-up* os DM são derivados do DW. Já na *top-down*, o DW é derivado dos DM. O *top-down* possui uma visão geral que vai do alto nível ao mais detalhado, ao passo que, o *bottom-up* vai do menor ao maior nível. Existe ainda uma terceira abordagem, abordagem híbrida, que constitui uma junção das duas metodologias anteriores (Kimball, 1998) (Inmon, 1996). Desta forma, o DW é indicado na construção de Sistemas de Apoio à Decisão. Nas secções seguintes é apresentado com mais detalhe as características associadas aos DW.

Kimball descreve no seu livro *Data Warehouse Toolkit* as principais características de um DW:

- Organização por assuntos: A informação é catalogada segundo a área de interesse da organização, assim a informação apresenta-se devidamente filtrada e classificada;
- Integrado: A integração dos dados obtidos das diversas fontes é feita dentro do DW;
- Não volátil: A informação presente no DW não se altera, apenas se incrementa;
- Variação no tempo: No DW a informação é armazenada por ordem cronológica;
- Metadados: São depósitos que armazenam dados sobre os próprios dados, por exemplo informação sobre tabelas da BD, número de colunas, tipos de dados, entre outros;
- Granularidade: É o nível de detalhe dentro do DW. Esta característica está relacionada com a conceptualização em diferentes níveis de abstração (Studer,

1998), (Mike Uschold, 1999), e portanto relacionada com o nível de detalhe ou precisão no qual o universo em causa está modelado.

Em relação aos DM, segundo (Alves, 2009), são nada mais que um subconjunto de informações existentes no DW, geralmente estruturados por assuntos. Quando um DW é composto por vários DM é sinal que a informação se encontra pensada e bem estruturada para satisfazer as regras do negócio. Esta decomposição em diversas áreas permite uma maior rapidez e eficiência na exploração dos dados. Quanto mais complexa for a sua estrutura mais complexa se torna a sua manutenção.



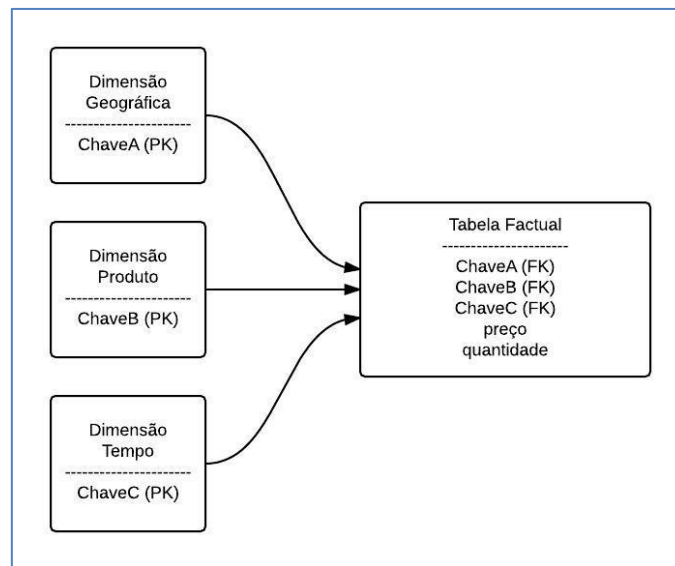
**Figura 5: Visão DW e DM**

As tabelas de factos, para (Geicer, 2003) são descritas como sendo as tabelas dentro de um modelo dimensional que contém as medidas e métricas de interesse para o negócio. Assim, a tabela de facto é constituída por valores numéricos relativos a um assunto específico do DW, a qual tem relacionamento com múltiplas tabelas de dimensão.

A tabela de factos possui a sua própria chave primária identificadora, que é formada por um subconjunto de chaves estrangeiras.

As tabelas de dimensão constituem pontos de entrada na tabela de factos. Atributos robustos nas dimensões oferecem capacidades analíticas robustas. As tabelas de dimensão implementam a interface do utilizador com o DW e representam frequentemente relações hierárquicas do negócio (Ralph Kimball, 2002). Por sua vez, as tabelas de dimensão incluem descritores e, contêm muitas colunas ou atributos. Nestas

tabelas é definida uma única chave primária. A interligação da tabela de facto e dimensão dão origem ao modelo dimensional.



**Figura 6: Componentes do Modelo Dimensional**

Fonte: Adaptado de Fidalgo, 2009

Em termos de medidas, ou métricas, são os atributos numéricos que representam um facto. Através delas é possível analisar o desempenho de um indicador de negócios relativo às dimensões que participam do facto (Kimball, 1998).

Como técnicas de modelação, a análise dos dados no DW é feita a partir de sistemas OLAP que permitem efetuar distintos tipos de consultas. Neste tipo de análise, denominada multidimensional, os dados são representados como dimensões em vez de tabelas. Nesta abordagem são consideradas estruturas de dados baseadas em dimensões e em factos. Segundo (Felber, 2005), existem várias técnicas de modelação multidimensional e relacional, dentre as quais, as mais utilizadas são: *Star Schema* (esquema estrela) e *Snowflake* ou *Star Schema* (esquema flocos de neve), os quais passamos a descrever:

- ***Star Schema***: Para (Adamson, 2010), este esquema sendo implementado numa BD relacional será apelidado de Esquema em Estrela, no entanto, implementado numa BD multidimensional será apelidado de cubo. O nome “Esquema em Estrela” surgiu devido à disposição dos componentes integrantes deste modelo se parecer com o formato de uma estrela (CF. Figura 7). O esquema em estrela agrega uma única

tabela de factos no centro e em redor múltiplas tabelas de dimensão. Cada dimensão está representada como uma única tabela. A chave primária em cada tabela de dimensão está relacionada com uma chave estrangeira na tabela de factos.

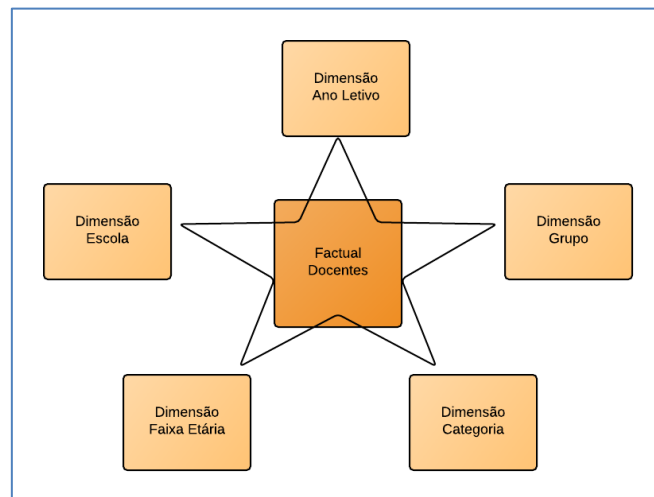


Figura 7: Modelo em Estrela

▪ **Snowflake**: Este modelo é considerado um refinamento do esquema em estrela onde parte da hierarquia dimensional é normalizada num conjunto de tabelas dimensão mais pequenas, numa forma similar a um floco de neve (Machado, 2004). A principal vantagem do esquema floco de neve é um melhor desempenho da consulta pois possui tabelas de pesquisa menores.

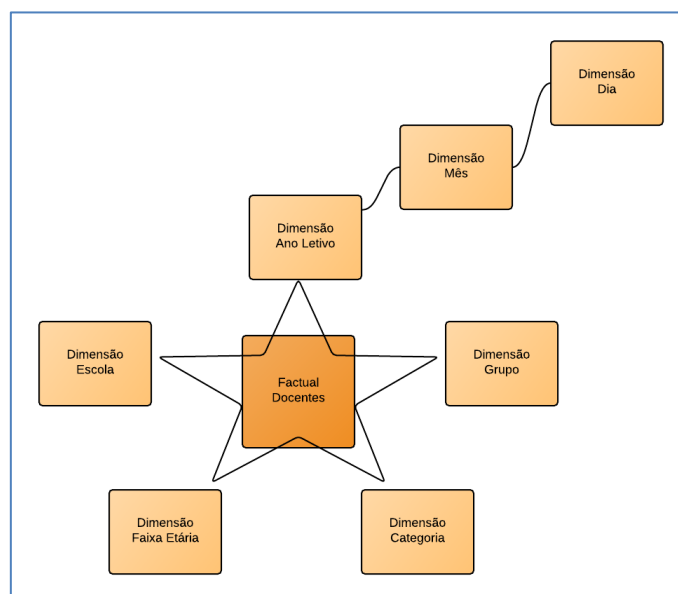


Figura 8: Modelo em Floco de Neve

A principal desvantagem deste esquema é o aumento da complexidade em manutenção devido ao aumento do número de tabelas de pesquisa. O esquema em floco de neve é semelhante ao esquema em estrela, diferenciando-se porque as tabelas dimensão encontram-se normalizadas num conjunto de tabelas de dimensão mais pequenas.

#### **2.2.4 PROCESSAMENTO OLAP – ON-LINE ANALYTICAL PROCESSING**

Após os dados serem modelados e carregados no DW, existem várias ferramentas de BI de exploração e visualização de dados que permitem obter e apresentar as informações que estão armazenadas no DW. Neste sentido, surgem a tecnologia OLAP. Os dados que são armazenados em bases de dados *Online Transaction Processing* (OLTP) do ponto de vista operacional são de enorme importância, porém, a forma como estão dispostos não ajuda na tomada de decisões. Neste sentido, surge a tecnologia OLAP para atender à necessidade de obtenção de informações que apoiem o processo de tomada de decisão. A tecnologia OLAP permite aos utilizadores acederem a dados organizados, agrupando subconjuntos de dados numa estrutura multidimensional que pode responder a consultas específicas. Em suma, esta tecnologia baseia-se na análise multidimensional dos dados e permite ao utilizador ter uma visão mais rápida e interativa dos mesmos. A grande vantagem das ferramentas OLAP é a comunicação entre a grande quantidade dos dados, permitindo ao utilizador (agente de decisão) ter uma visão analítica deles como um todo.

Assim, de seguida, apresentam-se as doze linhas de orientação descritas por (Codd, 1985):

- Ter uma visão conceptual multidimensional: facilitando a forma como os agentes de decisão contactam com os dados;
- Transparência: o acesso aos dados deve ser transparente para os utilizadores, ou seja, as funcionalidades das ferramentas de OLAP devem ser simples de utilizar;
- Facilidade de acesso: deve ser transparente o acesso às várias fontes de dados;
- Elaboração das consultas de análise devem ser consistentes: não deve ser dependente do número de dimensões;
- Arquitetura do OLAP deve ser do tipo cliente/servidor;

- Dimensionalidade genérica: para as ferramentas de OLAP, todas as dimensões são iguais e tem a mesma importância;
- Manipulação dinâmica de matrizes: os sistemas OLAP devem ser capazes de se adaptarem ao seu esquema físico, mantendo os elevados requisitos de desempenho;
- Devem suportar vários utilizadores;
- Operações entre dimensões sem restrições: os sistemas OLAP devem ser capazes de reconhecer hierarquias dimensionais e automaticamente efetuar os cálculos associados às dimensões;
- Manipulação de dados intuitiva: permite a análise através de análises de agregação *Roll-up* e de refinamento *Drill-down*;
- Devem permitir apresentar relatórios flexíveis às análises complexas invocadas;
- Níveis de agregação e dimensões ilimitados: as ferramentas devem ser capazes de agregar muitas dimensões.

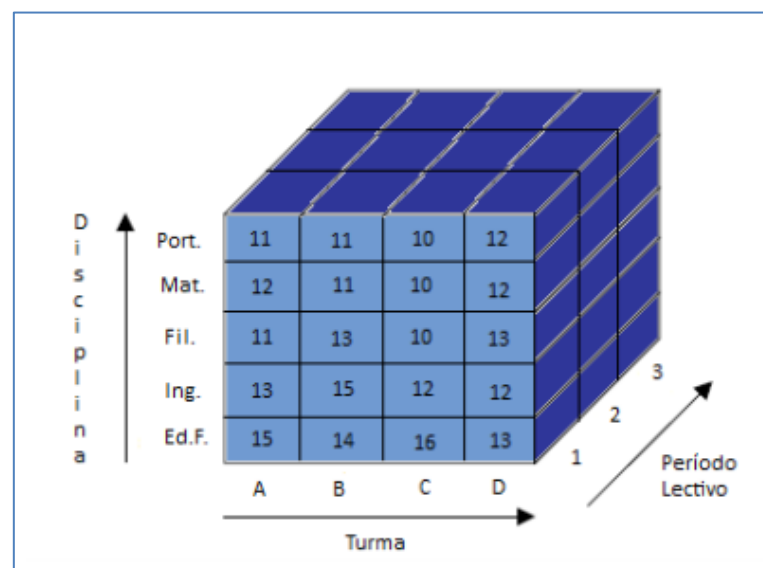
As ferramentas OLAP podem ser caracterizadas em três categorias. MOLAP (*Multidimensional OnLine Analytical Processing*), ROLAP (*Relational OnLine Analytical Processing*) e MQE (*Managed Query Environment*). Os sistemas MOLAP, são caracterizados por processarem informação que já está armazenada numa estrutura multidimensional, onde todas as combinações possíveis de informação são refletidas, sendo cada célula acedida diretamente. Por este facto para muitos investigadores o acesso à informação pretendida é mais rápida através de MOLAP do que o ROLAP. O ROLAP permite rentabilizar as múltiplas análises multidimensionais dos dados armazenados. As análises pretendidas pelos utilizadores são convertidas em comandos SQL que são submetidos à base de dados. A resposta dos motores de ROLAP, é uma conversão do resultado do comando SQL num formato multidimensional antes de ser apresentado ao utilizador. Tipicamente como se trata-se de bases de dados relacionais, requerendo bastante tempo de processamento e espaço em disco para executar determinadas *queries* complexas.

As ferramentas MQE possuem a capacidade de oferecer análises "*datacube*" e "*slice and dice*". Primeiro há a necessidade de efetuar uma consulta aos sistemas de gestão de bases de dados (SGBD) para selecionar a informação, sendo de seguida guardados num

cubo de dados local, onde os utilizadores podem efetuar localmente análises multidimensionais.

Em suma, OLAP é uma tecnologia que foi introduzida em 1990 possibilitando a análise de dados provenientes de aplicações baseadas em transações (Shoshani, 1997). Considerada uma ferramenta de BI, pode ser definida como um estudo multidimensional e interativo da informação de uma instituição. Conforme afirmam (Araújo, Batista, & Magalhães) a tecnologia OLAP possibilita às organizações um método de acesso, visualização, e análise de dados corporativos com alta flexibilidade e desempenho, por meio de relatórios e análises a cubos de dados.

Os cubos de dados OLAP ou estruturas multidimensionais são um tipo de matriz multidimensional que permite que os utilizadores explorem e analisem uma coleção de dados a partir de várias perspetivas diferentes, geralmente considerando três fatores (dimensões) de cada vez (Kay, 2010). Assim sendo, OLAP permite valer-se da informação desde diferentes perspetivas, perspetivas essas apelidadas de dimensões. Como pode ser visto na seguinte figura, o cubo é a estrutura multidimensional que utiliza OLAP para albergar a informação.



**Figura 9: Cubo Genérico**

Fonte: Trabalho de Mestrado

Desenvolvimento de Software e Sistemas Interativos (2011)

Na figura anterior é visível que o cubo contém as dimensões de Disciplina, Turma e Período Letivo. As células correspondentes à interseção das dimensões contêm valores de medidas.

Harmonizar os dados num cubo permite maior flexibilidade e evita as limitações das BD relacionais que não são tão eficientes na análise de grandes volumes de dados. Um cubo de dados é constituído por tabelas de factos e dimensões.

A partir da estrutura multidimensional do cubo é possível aplicar várias operações OLAP para análise de dados.

- Slice - Executa uma seleção sobre uma das dimensões de um cubo, resultando num sub-cubo (ou sub estrutura multidimensional);
- Dice - Extrai um sub-cubo do cubo original executando uma operação de seleção em duas ou mais dimensões;
- Drill Down - Aumenta o nível de detalhe da informação consultada;
- Roll Up - Diminui o nível de detalhe da informação consultada;
- Drill Across – Permite aceder a dados que estão em outros cubos;
- Drill Through – Permite aceder a dados que se encontram no DW;
- Pivot – Permite rodar o cubo para poder observá-lo em diferentes planos, adicionar ou readaptar as dimensões;
- Rotation – Permite alterar a ordem das dimensões e visualizar os dados numa nova perspetiva.

Em termos de esquema para estruturação do cubo, o *schema* é escrito em linguagem XML e pode ser criado num editor de texto. Cada *schema* tem um nome, um cubo (coleção de medidas e dimensões), uma tabela de factos, dimensões e medidas.

Cada medida tem um nome, uma coluna na tabela de factos e um agregador. O agregador é por exemplo o operador "sum" ou "max".

Abaixo pode ser visto um exemplo da estrutura de um cubo:

```
<Schema name="InfoPintor">
  <Cube name="Funcionario">
    <Table name="facto_funcionario"/>
    ...
    <Dimension name="dim_categoria"
      foreignKey="id_categoria">
      ...
```



```
        <Measure aggregator="distinct-count" caption="N_
        Alunos" column="num_aluno" name="numero_alunos"
        </Measure>
    </Cube>
</Schema>
```

### 2.2.5 VISUALIZAÇÃO DE DADOS -DASHBOARDS

*“Um Dashboard é uma representação visual das informações mais importantes necessárias para atingir um ou mais objetivos, consolidada e organizada num único ecrã de modo a que a informação possa ser monitorizada rapidamente” (Few, 2004).*

Tal como descrito anteriormente, um *Dashboard* possibilita a demonstração de indicadores e a publicação de relatórios interativos. Existem dois tipos de *Dashboards*:

- Analítico: permitem obter, a partir de DM, relatórios e indicadores chave de desempenho (*Key Performance Indicators* (KPIs)), análise de áreas de negócios não relacionadas entre si. Na prática, é uma ferramenta de consulta que visa a obtenção e apresentação de indicadores de gestão;
- Integrais: são desenvolvidos estrategicamente ao nível de toda a organização. Permitem aos diferentes níveis de gestão e liderança da organização ter uma visão estratégica dos indicadores, uma vez que,
  - cobrem toda a organização.

Através das diferentes ferramentas de representação e visualização (OLAP, DM, e *Dashboards*), o agente de decisão tem acesso ao conhecimento obtido e, com base nas informações disponibilizadas, poderá proceder a uma tomada de decisão baseada em evidências. Neste sentido, o *Dashboard* pode ser interpretado como um veículo de comunicação, essencialmente de visualização gráfica, onde se expõe um conjunto de informação muito relevante. Os destinatários podem ser, na prática, qualquer pessoa em qualquer ponto da organização, ou mesmo fora dela – gestores de topo, intermédios, colaboradores e *stakeholders* (Caldeira, 2010). Trata-se de um conceito simples, porém poderoso, que tem criado imenso impacto em diversas áreas da engenharia e ciência.

Cada vez mais existe a necessidade de sistemas que permitam visualizar informação pertinente, credível e bem estruturada, de modo a permitir responder de forma imediata

e fundamentada, às principais questões colocadas pelos responsáveis por tomar decisões (Caldeira, 2010).



**Figura 10: Exemplo Dashboard**

Fonte: Google Analytics

## 2.3 FERRAMENTAS/PLATAFORMAS DE BUSINESS INTELLIGENCE

Cada vez mais, são procuradas aplicações informáticas de forma a suportar as atividades operacionais das organizações. Por outro lado, os *softwares OS* consistem em modelos de desenvolvimento no qual o código fonte está inteiramente disponível aos utilizadores para visualização, leitura, modificação e redistribuição sem as restrições de direito de propriedade do produto. O *software OS* difere do não *OS* nos processos de desenvolvimento e nas licenças do produto. Todas as aplicações de código aberto encontram-se licenciadas por uma licença *OS* que proporciona ao utilizador o direito de utilização do *software*, acesso e modificação do seu código fonte e a redistribuição desse mesmo (Maddocks & Waring, 2005).

Não se deve descurar a importância da escolha adequada de uma ferramenta de BI. Uma vez que existe uma grande variedade no mercado, o processo de seleção não é trivial. É importante definir desde cedo o que se pretende com a aplicação de BI, antecipar-se às necessidades futuras e não esquecer os utilizadores que vão estar em contacto direto com a ferramenta.

De forma a perceber qual a ferramenta de BI *OS* que melhor se enquadrava no âmbito da educação, foi realizada uma revisão da literatura sobre qual seria a melhor ferramenta

OS que se aplicaria aos requisitos pretendidos pelo Agrupamento. Na área da educação, os requisitos para implementação de um sistema de BI passam essencialmente por obter informação que ajude na tomada de decisão a um nível estratégico, operacional, utilizando dados históricos e atuais, e visualizando-os através de relatórios ou gráficos.

Por sua vez, a literatura apresenta diversos trabalhos comparativos entre as diversas plataformas de BI. Tereso e Bernardino, em 2011, criaram um quadro que apresenta uma comparação entre ferramentas de BI, debruçando-se mais especificamente sobre as OS. Baseado em revisões de literatura e pesquisas acerca das ferramentas mais conhecidas no mercado, a tabela a seguir apresentada resulta de uma análise relativa ao estado da arte de ferramentas de BI.

	Solução completa	Relatórios	Gráficos	OLAP	ETL	Dashboard	Open Source
 Microsoft Business Intelligence	X	X	X	X	X	X	
 ORACLE BUSINESS INTELLIGENCE	X	X	X	X	X	X	
 SAP BI	X	X	X	X	X	X	
 pentaho	X	X	X	X	X	X	X
 openi.org Open Intelligence		X	X	X		X	X
 SpagoBI	X	X	X	X	X	X	X
 TARGIT business intelligence	X	X	X	X	X	X	

Tabela 1: Comparação entre Ferramentas de BI

Como pode ser observado a *suite* Pentaho oferece um leque variado e completo de funcionalidades. Esta *suite* é apresentada em duas versões: a versão *Community* em que parte das funcionalidades são gratuitas e a versão *Enterprise* que é paga e oferece todas as funcionalidades BI, bem como suporte e formação (Golfarelli, 2009).

Todas as ferramentas de BI anteriormente apresentadas são muito semelhantes em termos das funcionalidades disponibilizadas. De uma forma geral, e observando a tabela anterior, tendo em conta as características avaliadas e importantes no âmbito deste projeto, as ferramentas Pentaho CE e SpagoBI são as mais completas (Lapa, Bernardino, & Figueiredo, 2014). Porém, após análise das ferramentas selecionadas e instalação das mesmas, constatou-se que a ferramenta SpagoBI, apesar de ser muito completa e totalmente gratuita, é de difícil instalação. Neste sentido, inferiu-se que o Pentaho CE, seria uma boa opção para a construção da plataforma de BI pretendida. Esta é de fácil instalação e utilização, permitindo realizar com sucesso muitas das tarefas de BI que atualmente várias organizações necessitam e concretamente permite realizar muitos dos requisitos impostos neste projeto. Neste sentido, na próxima secção será abordado o Pentaho CE, de forma a avaliar e a analisar mais especificamente a ferramenta em questão.

### 2.3.1 PENTAHO COMMUNITY EDITION

O *software* foi desenvolvido pela Pentaho Corporation, em 2001, na linguagem Java, sendo a primeira plataforma de BI OS a ser lançada no mercado. O Pentaho tem compatibilidade técnica com os principais sistemas operativos (Windows, Linux ou Mac) e possibilita a ligação às seguintes bases de dados de JDBC: IBM DB2, Microsoft SQL Server, MySQL, Oracle, PostgreSQL, Firebird entre outras.

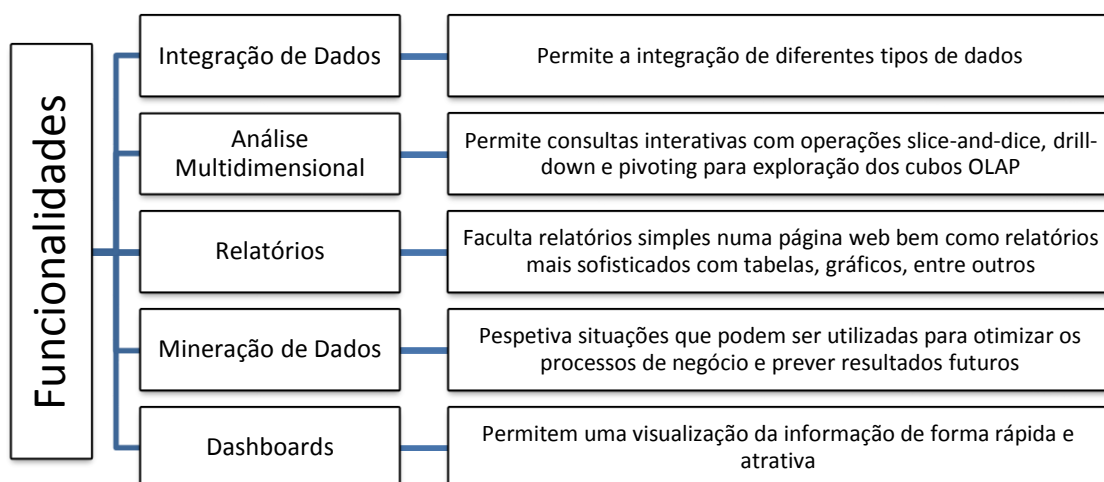
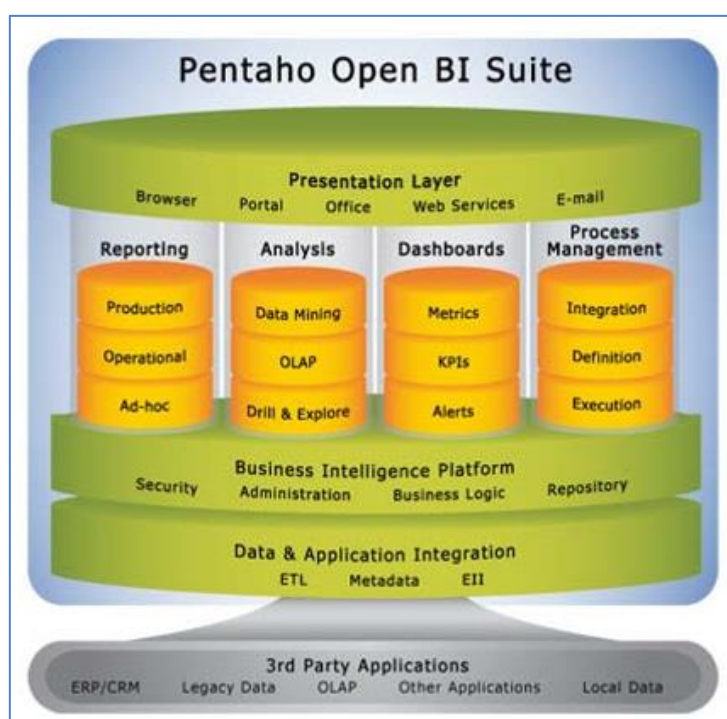


Figura 11: Pentaho BI Suite

O Pentaho BI Suite como é apelidado pelo projeto da Pentaho para BI, é uma ferramenta completa para análise de dados multidimensional. Incorpora o Pentaho BI Platform (Bi-Server); a ferramenta Kettle para abordagens de ETL (também designada de Pentaho Data Integration); para o cruzamento de dados OLAP incorpora a ferramenta Mondrian; possui funcionalidades próprias para criação de relatórios de dados; integra também algoritmos sofisticados de mineração de dados baseadas na aplicação Weka e disponibiliza ainda propriedades relacionadas com a criação e visualização de *Dashboard*. Por sua vez, o Pentaho permite o *download* da solução BI completa ou unicamente das ferramentas específicas de que o utilizador precise. A figura seguinte ilustra o processo da plataforma do *Pentaho Open BI Suite*.



**Figura 12: Plataforma Pentaho Open BI Suite**

Fonte: Miranda, 2010

A *suite* Pentaho é constituída por vários componentes *open source* com funcionalidades muito abrangentes e bastante completa para muitas organizações, sendo constituída pelo Mondrian – Servidor OLAP onde são explorados os cubos de informação; pelo Schema Workbench – Permite a criação dos cubos; pelo Kettle – Ferramenta ETL de Integração de dados; pelo Saiku – Visualização de cubos multidimensionais; pelo JFreeReport – Permite gerar relatórios e pela ferramenta Weka – Mineração de dados.

## 2.4 METODOLOGIAS DE IMPLEMENTAÇÃO DE PLATAFORMAS DE BI

Nesta secção serão abordadas as metodologias técnicas para a construção de um DW, em particular a metodologia de Ralph Kimball e a de William Inmon. Existem várias metodologias técnicas e de desenvolvimento de um projeto de DW. As mais conhecidas são a abordagem de Inmon e a abordagem de Kimball. Segundo a metodologia de Kimball, esta começa com um DM e posteriormente são adicionados mais DM, caso necessário. Desta forma o fluxo de dados tem origem nos DM e em seguida, no DW, sendo por isso mais rápida. Ao passo que, na abordagem de Inmon, o DW é desenvolvido com base no modelo de dados de toda a instituição, sendo por isso mais lenta a sua implementação. Neste sentido, a metodologia seguida neste projeto foi a de Kimball. As fases desta metodologia são apresentadas na figura seguinte (Kimball, 1998).

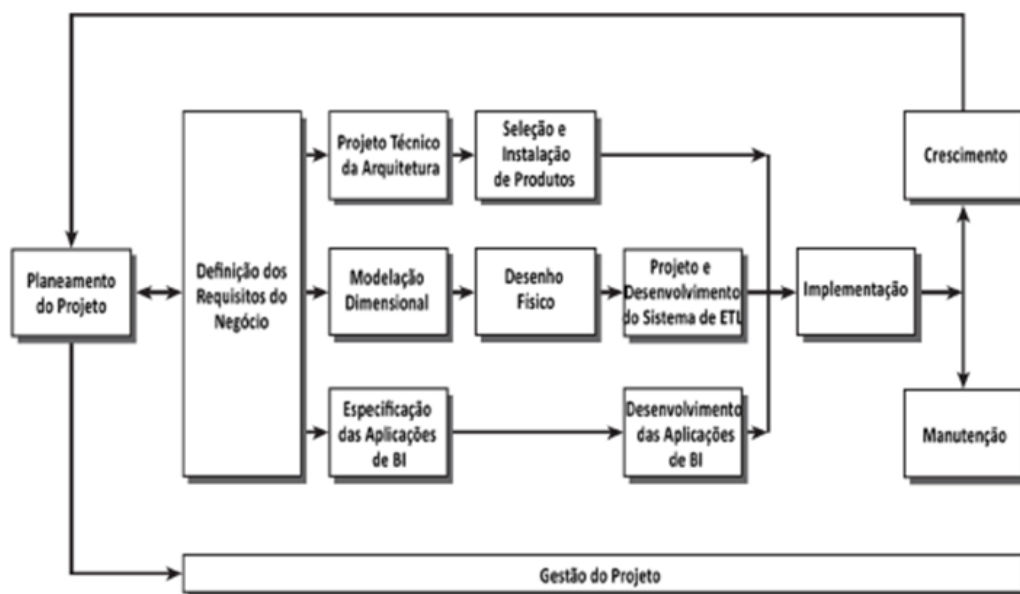


Figura 13: Ciclo de Vida da metodologia de Kimball

Fonte: Kimball, 1998

O ciclo de vida do desenvolvimento inicia-se com a fase de planeamento do projeto que segundo Kimball pode variar de empresa para empresa. Esta fase é caracterizada por uma atividade mais crítica, pois a qualidade das informações recolhidas e definições gerais podem afetar o projeto como um todo.

Na fase de definição dos requisitos do negócio são definidos os fatores chave do projeto, imprescindíveis para as próximas etapas. No diagrama da figura anterior pode-se

observar três sequências de tarefas que ocorrem em simultâneo. Estas estão focadas na escolha da tecnologia, na modelação e no desenvolvimento da aplicação de BI. A fase tecnológica inclui:

- Projeto técnico da arquitetura: segundo Kimball, esta fase é uma das mais importantes no projeto de DW. É definida uma arquitetura de alto nível (área interna e área externa) e a especificação da infraestrutura técnica e os respetivos componentes necessários para permitir a criação do DW;
- Seleção e Instalação dos Produtos: por outro lado, esta fase consiste na realização de tarefas de investigação, estudo e seleção dos produtos relacionados para a construção de um DW. Deve também ser desenvolvido um protótipo para ser possível avaliar melhor as funcionalidades dos *softwares* e a aceitação destas pelo agente de decisão.

A modelação inclui, a Modelação dimensional: consiste em agregar os dados levantados na fase de definição dos requisitos do negócio para desenvolver um modelo dimensional. Existem várias abordagens de modelagem dimensional, sendo as mais conhecidas o esquema em estrela e o esquema em floco de neve; o Desenho físico: a execução do desenho físico consiste na criação dos objetos da base de dados analítica e no desenvolvimento de um plano inicial de indexação e agregação; e o Projeto e desenvolvimento do sistema de ETL: Esta fase diz respeito às atividades fundamentais a serem realizadas no DW: a extração, transformação e carregamento dos dados.

O Desenvolvimento aplicacional compreende, compreender a especificação das aplicações de BI; o desenvolvimento de aplicações de BI: são desenvolvidas as aplicações necessárias de acordo com os levantamentos realizados na fase de especificação de aplicações de agentes de decisão. A seleção do ambiente de desenvolvimento dos relatórios, o desenvolvimento de procedimentos de manutenção e atualização das aplicações para o utilizador final são atividades que compreendem esta fase.

A fase de manutenção e crescimento do DW é composta basicamente pelo contínuo suporte e treino dos utilizadores e manutenção da infraestrutura técnica, além da monitorização de consultas realizadas pelos utilizadores, desempenho da organização de dados e o contínuo sucesso do DW.

## **2.5 CONSIDERAÇÕES FINAIS**

Este capítulo abordou os principais conceitos sobre BI que foram importantes para o desenvolvimento da solução proposta neste trabalho. Não foi explorado a parte de *Data Mining* e consequentemente nenhuma metodologia orientada para este tipo de implementação, podendo a mesma ser explorada em trabalhos futuros.



## 3. CASO DE ESTUDO - ANÁLISE DE REQUISITOS

### 3.1 INTRODUÇÃO

Para ajudar a compreender as oportunidades que a informação fornece e promover análises exploratórias avançadas, as Escolas devem procurar novas tecnologias de visualização objetiva. Dotadas destas ferramentas, as instituições de ensino, conseguem extrair informação crucial e relevante, facilitando assim as tomadas de decisão e agindo de forma rápida e precisa. Além disso, com o apoio da visualização da informação, os gestores podem resolver problemas complexos que seriam impossíveis sem o apoio visual (Speier, Vessey, & Valacich, 2003).

Este capítulo contempla todo o processo de recolha de informação por parte dos elementos estruturantes da escola. Baseado no modelo clássico de BI, a aplicação produzida foi pensada e estruturada em partes: o processo ETL, o desenvolvimento do DW, a análise da informação recorrendo a ferramentas OLAP e o desenvolvimento de uma plataforma web com *Dashboards* de apoio e visualização de informação ao utilizador.

### 3.2 A INSTITUIÇÃO DE ENSINO ALVO DO ESTUDO

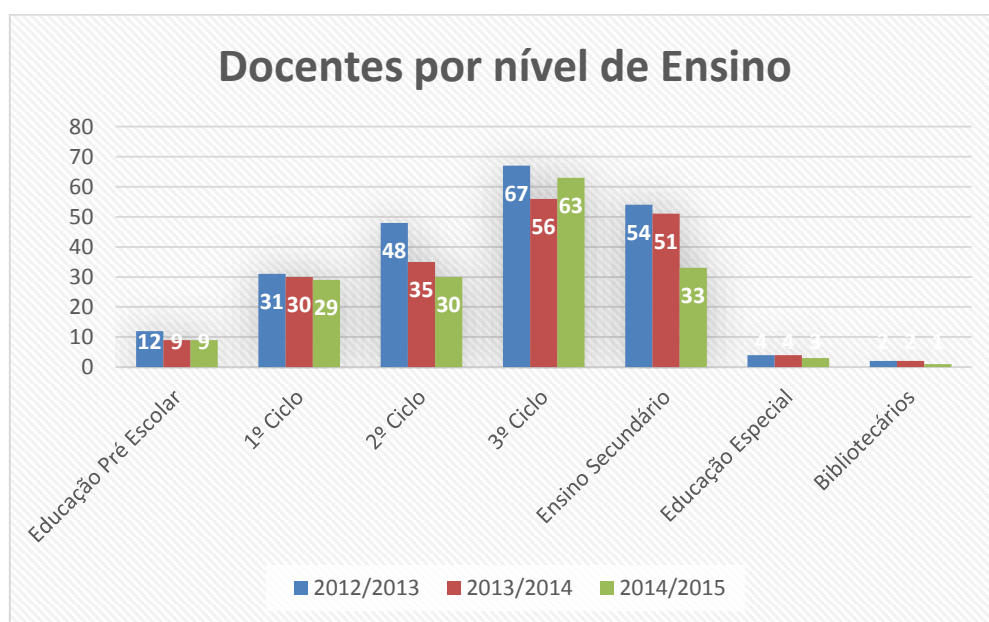
O APJB viu homologada a sua constituição pelo Despacho de 19/04/2002, ao abrigo do artigo 8.º do Decreto-Lei n.º 115- A/98, de 4 de maio, e do Decreto Regulamentar n.º 12/2000, de 29 de agosto. A sua designação, recuperando o nome da escola que viria a ser a sede do Agrupamento, constitui uma homenagem a José de Brito (1855-1946), pintor oriundo da freguesia de Santa de Marta de Portuzelo, concelho de Viana do Castelo, que também foi professor na Academia Portuense de Belas Artes, no início do século XX. Atualmente, e de acordo com a Portaria n.º 30/2014, de 5 de fevereiro, o APJB é constituído pelos seguintes estabelecimentos de educação e ensino:

- Escola Básica e Secundária Pintor José de Brito, Santa Marta de Portuzelo (Escola Sede)
- Escola Básica de Portuzelo, Meadela
- Escola Básica de Igreja, Cardielos
- Escola Básica de Igreja, Nogueira

- Escola Básica de Moreno
- Escola Básica de Outeiro
- Escola Básica de Santa Marta de Portuzelo
- Escola Básica de Perre

O APJB reparte a sua oferta formativa desde o pré-escolar até ao 12º ano. A nível secundário oferece os cursos Científico-Humanísticos, Profissionais, CEF's e Vocacionais.

No ano letivo 2014/2015, o APJB contou com 1149 alunos distribuídos por 57 turmas, 136 docentes, 46 funcionários e 1 psicóloga. Os docentes, no geral são maioritariamente do quadro de escola, com mais de 17 anos de serviço e predominantemente do sexo feminino. O número de pessoal não docente (assistentes operacionais e assistentes administrativos) tem-se mantido estável ao longo dos últimos anos. No gráfico seguinte ilustram-se, alguns dados mais específicos sobre a população escolar, procurando uma caracterização um pouco mais específica do Agrupamento.



**Figura 14: Distribuição de docentes por nível de ensino**

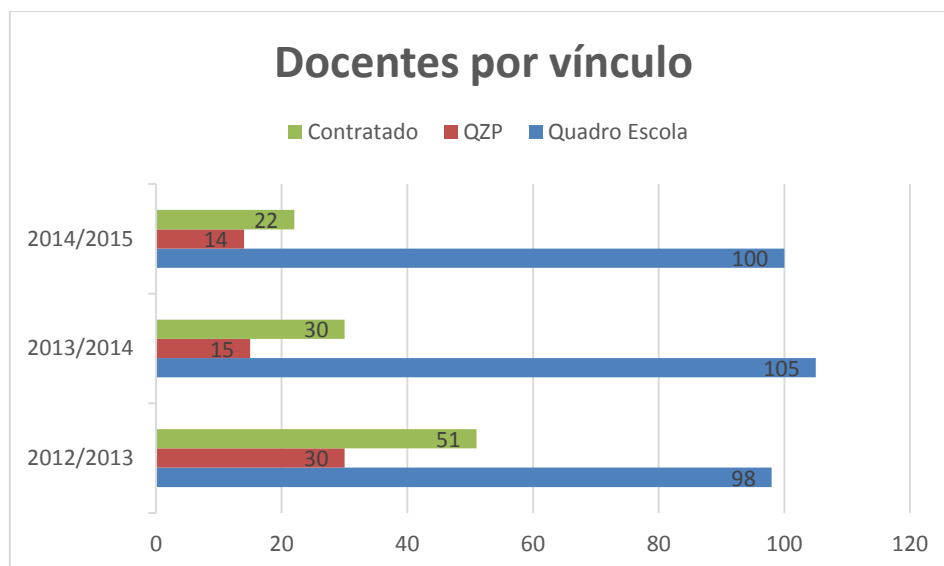


Figura 15: Distribuição de docentes por vínculo

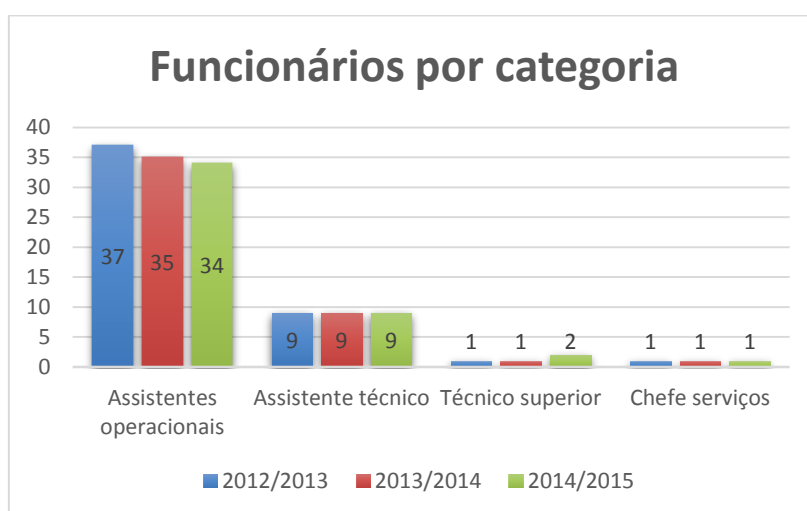


Figura 16: Distribuição de funcionários por categoria

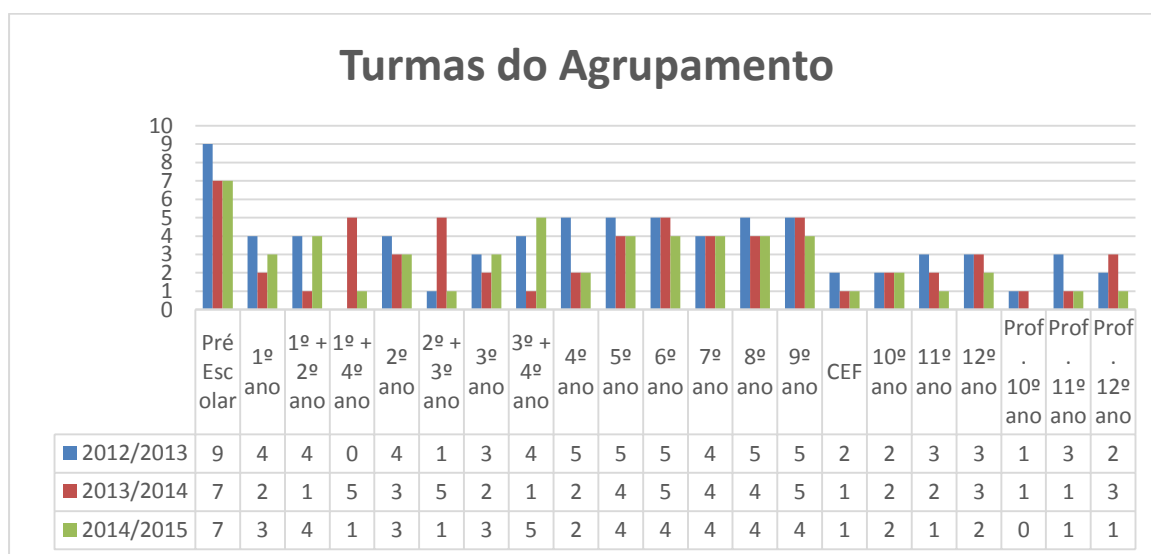
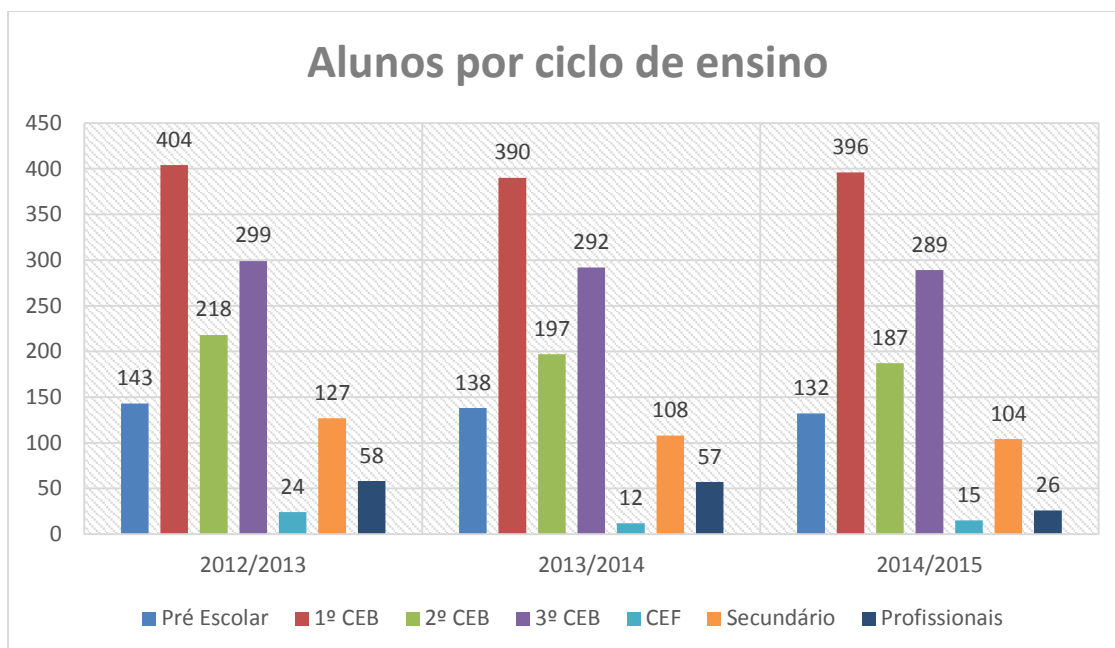


Figura 17: Distribuição das Turmas do Agrupamento por ano



**Figura 18: Distribuição do nº de alunos por ciclo de ensino**

No APJB, um estabelecimento do ensino básico e secundário, constatou-se um problema que consiste em não conseguir compilar, em tempo, útil, as informações disponíveis nas várias plataformas usadas (Aplicação Alunos, Aplicação Multiusos, GPV, ENES, ENEB, PFEB, entre outros). Estas aplicações, constituídas por soluções que são da responsabilidade direta da Escola, e outras disponibilizadas pelo próprio Ministério da Educação, geram grandes quantidades de dados complexos, falta de informações ou informações às vezes inconsistentes ou redundantes. Este problema, associado à quantidade de informação a tratar, contribui para a existência de erros na tomada das decisões por parte dos gestores, ao mesmo tempo que as tornam menos eficazes (Pires, 2011); (Santos A. P., 2013).

No passado, o principal problema da procura de informação era a recolha de dados. Atualmente, o desafio não é a recolha de dados, mas a sua localização e interpretação para extração de informações (Santos M. J., 2011).

Nos últimos anos, temos assistido a um crescimento explosivo dos dados ligados à educação. Os Diretores e a sua equipa precisam de sistemas eficazes devido à crescente complexidade da gestão de dados institucionais. Na linha de pensamento de (Yusuf, 1998), essa complexidade é provocada pelo:

- aumento da população escolar;
- aumento de programas e recursos;

- aumento das pressões sobre as instituições educativas para responder às necessidades e expectativas da sociedade;
- aumento da prestação de contas do sistema educativo e da sua gestão;
- aumento de conhecimento e informação dentro e fora da instituição.

A razão principal que leva à utilização de aplicações de gestão (alunos, funcionários, vencimentos, avaliações, entre outras) por parte do Agrupamento é a de melhorar a eficiência das atividades de gestão escolar. Estas aplicações, por exemplo, a aplicação Alunos, é utilizada para registar todos os dados pessoais relativos a estes alunos bem como o seu percurso escolar. A aplicação GPV é utilizada para registar dados relativos aos professores e funcionários que trabalham no Agrupamento. (Shah, 2014) afirma que a gestão do sistema de informação tem um impacto positivo na administração e gestão escolar, favorecendo também acessibilidade à informação, uma administração mais eficiente, maior utilização dos recursos educativos, redução da carga de trabalho, melhoria da gestão do tempo e na qualidade dos relatórios.

(Daradkeh, 2012) destaca que o sistema de informação fornece aos gestores as informações necessárias para o planeamento de objetivos e a formulação de políticas. Ressalta a sua importância nas áreas de liderança, na tomada de decisão, na gestão de recursos humanos, comunicação e responsabilidade. Conclui que os gestores têm dificuldade no tratamento das informações e na tomada de decisões corretas, por isso precisam fazer uso de ferramentas para auxiliar no tratamento e análise dos dados.

Uma ferramenta cada vez mais usada por parte das chefias para manipulação de dados e tomada de decisão, são os *Dashboards*.

### **3.3 TRABALHOS RELACIONADOS**

Em termos de trabalhos relacionados com a aplicabilidade dos sistemas de suporte à decisão, em particular nas ferramentas de BI, realizamos pesquisas sobre trabalhos relacionados com este projeto, nomeadamente ao nível da utilização dos sistemas de BI no ensino, e que permitiu concluir que já existem alguns trabalhos desenvolvidos nesta área. Em Portugal existem vários setores que utilizam sistemas de BI: governo, câmaras municipais, finanças, educação e saúde.

A dissertação de mestrado apresentada por (Diamantino, 2012), ao Instituto Politécnico de Castelo Branco, aborda a implementação das tecnologias DW e BI, no Agrupamento Afonso de Paiva em Castelo Branco. O trabalho assenta numa aplicação desenvolvida para acompanhamento escolar, com o objetivo de permitir traçar metas educativas para os alunos.

O trabalho publicado por José Guerra e (Guerra & Mendes, 2010), permite-nos verificar como foi desenvolvido e implementado um Sistema de Apoio à Decisão Escolar, na Escola Secundária das Laranjeiras, em Ponta Delgada. Este sistema, bastante completo, possibilita a análise da informação de várias escolas e mesmo de Agrupamentos.

Como exemplos concretos temos os projetos desenvolvidos para a Câmara Municipal de Vila Verde, para a Escola Secundária/3 de Amato Lusitano, para a Universidade de Coimbra, entre outros.

A pesquisa bibliográfica envolveu também o estudo de ferramentas OS que proporcionam a implementação e desenvolvimento do DW. Como abordado anteriormente, foi feito um comparativo das principais ferramentas, contribuindo assim para uma escolha mais adequada da ferramenta a usar neste projeto (CF. Tabela 1).

Numa outra vertente de análise de trabalho relacionado e mais concretamente em termos de *Dashboards*, foi feita uma análise comparativa de abordagens existentes e das suas vantagens e inconvenientes. No Excel, existem vários recursos que podem ser combinados a fim de se obter um *Dashboard* de alto nível, oferecendo um *layout* agradável e bastante intuitivo para o utilizador. Conscientes da importância dos *Dashboards* como uma realidade cada vez mais presente nas instituições, estas recorrem ao Excel como uma poderosa ferramenta de trabalho. Muito rapidamente conseguem criar painéis de comparação com o objetivo de dar uma visão rápida e clara da situação atual de um determinado setor da instituição.

A Google tem uma aplicação, *Google Analytics*, que consiste numa solução de análise da web a nível empresarial e que proporciona a monitorização de informações valiosas (nº de visitantes, taxa de rejeição, páginas visitadas, entre outros) sobre o tráfego do website de qualquer organização e assim perceber a eficácia das políticas de uma instituição. O *Google Analytics* é uma ferramenta de web *analytics* gratuita, que permite monitorizar a origem e o comportamento dos utilizadores de um website.

A *Sprout Social* é uma ferramenta de gestão de redes sociais. Esta aplicação permite obter rapidamente informações genéricas de uma página, o crescimento do número de fãs e o desempenho do conteúdo do Facebook, com relatórios intuitivos e interativos.

A *icCube Web Reporting* é uma ferramenta que permite criar *Dashboards* interativos através de um *Web Browser*. Utilizando esta aplicação é possível adicionar componentes gráficos aos *Dashboards*, bem como ativar operações de *drill-down* sobre os dados e permitir também a aplicação de filtros.

Muitas outras ferramentas análogas existem no mercado, como por exemplo a *Dundas Dashboard* e a *InetSoft Style Scope*, entre outras, que apresentam inúmeras características interessantes e serviços para a criação de sistemas de *Dashboarding*. Estas ferramentas oferecem um elevado grau de liberdade na configuração dos componentes gráficos, sendo extremamente completas mesmo quando se pretende efetuar níveis elevados de customização.

Segundo os dados recolhidos sobre as características gerais dos programas selecionados, quanto à licença, observa-se que todos apresentam uma versão experimental de 30 dias e o *Google Analytics* é totalmente gratuito. Em relação à plataforma, a maioria disponibiliza a aplicação *online*. Quanto aos sistemas operativos, todos os programas possuem versão para sistemas Windows, Linux e Mac OS X. Observa-se também que todas as aplicações possuem um leque variado de formatos de exportação de ficheiros.

Em Janeiro de 2015, o Ministério da Educação e Ciência lançou publicamente o portal InfoEscolas (<http://infoescolas.mec.pt>). Este portal divulga informação estatística sobre as escolas secundárias (cursos científico-humanísticos) sediadas em Portugal Continental.

Este portal é, sem dúvida, um instrumento de trabalho útil para os agentes de decisão de um Agrupamento e para os seus professores, no entanto mostra-se deficitário pois não apresenta todos os cursos, anos letivos e não permite a seleção das escolas de um Agrupamento. É uma aplicação pouco flexível, com um espectro de dados limitado e sem possibilidade de *drill*.

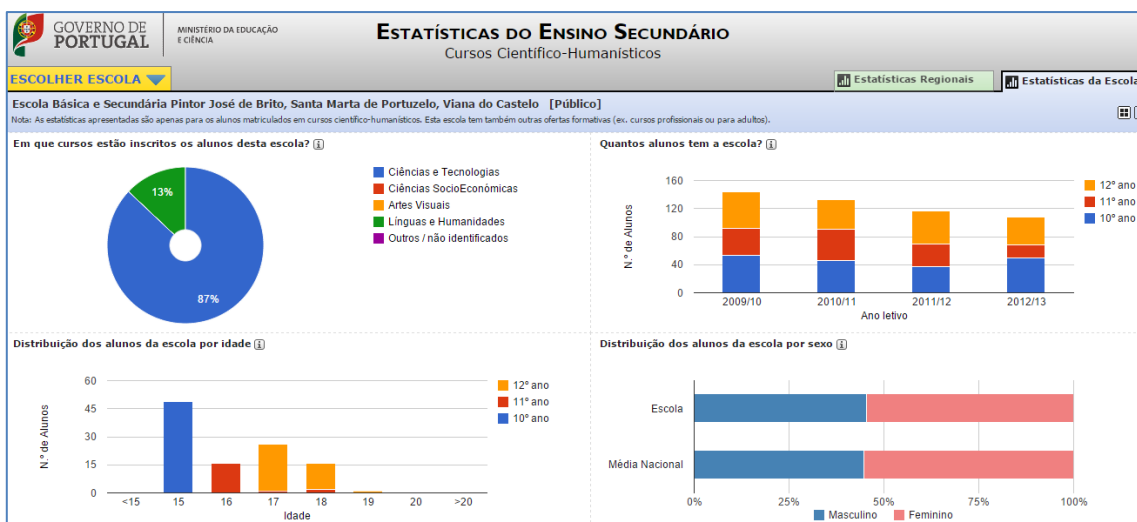


Figura 19: *Dashboard* Educacional - Infoescolas

Pela nossa análise e consulta, constatamos, nesta altura que as aplicações disponíveis no mercado não respondem às especificações pretendidas, excluem certos cursos (profissionais e vocacionais) e não permitem visualizar dados relevantes sobre professores, alunos e funcionários.

### 3.4 REQUISITOS

Na primeira fase da metodologia de Kimball (definição do problema) definiu-se o âmbito do projeto como sendo o desenvolvimento de uma plataforma de BI que permitisse a obtenção de indicadores de gestão. Observou-se que os registos armazenados não eram utilizados, ou quando utilizados, os agentes de decisão das escolas recorrem a técnicas rudimentares, demasiado demoradas e dispendiosas (por exemplo utilizando a folha de cálculo). Por sua vez, realça-se a necessidade e atenção no cuidado com as questões como a qualidade da informação, pois têm implicações diretas no processo de uma tomada de decisão eficaz. Neste sentido a principal motivação passa pelo desenvolvimento de uma plataforma de BI capaz de permitir a obtenção de indicadores de monitorização.

De acordo com os aspetos referidos anteriormente, definiram-se e planearam-se as atividades a executar ao longo de todo o projeto. Esse planeamento iniciou-se com a análise das informações necessárias para o caso de estudo. Nesta etapa:

- planearam-se a marcação de reuniões com os agentes de decisão inseridos na APJB com o intuito de perceber quais os objetivos da plataforma, requisitos e quais os indicadores a obter no final;



- foram analisados os dados disponíveis: validação e importação dos dados para o Sistemas de Gestão de Bases de Dados (SGBD) - nesta etapa devem ser tidas em conta as questões de ETL;
- foram construídos mecanismos para automatizar o processo: nesta etapa foi estudada a melhor forma para a importação dos dados de acordo com a periodicidade (semanalmente, anualmente, etc.) e foi automatizado o processo de ETL;
- foram criados os cubos ou *queries*: após o carregamento dos dados na estrutura multidimensional (DW), foram criados os cubos e as *queries* necessárias para a criação dos indicadores pretendidos;
- procedeu-se à configuração das permissões de acesso: numa fase final devem ser criados grupos de utilizadores com diferentes permissões de acesso de acordo com os requisitos impostos pelos agentes de decisão.

A fase de compilação dos requisitos é considerada por alguns autores como a parte mais crítica no desenvolvimento de *software*, pois a qualidade do produto final depende fortemente da qualidade dos requisitos (Ferguson & Lami, 2006). Pesquisas apontam que 85% dos problemas de *software*, têm origem na atividade de recolha de requisitos (Fernandes, Machado, & Seidman, 2009).

A fase de especificação dos requisitos teve início em várias reuniões com elementos da direção e com os responsáveis da secretaria pelas áreas de alunos e docentes da escola, com o objetivo de perceber o intuito geral de toda a aplicação. Estas reuniões foram fundamentais para perceber a dinâmica da escola e compreender as reais necessidades a explorar. Durante a especificação dos requisitos, foram identificadas quatro áreas prioritárias: Alunos, Professores, Funcionários e Escolas, estando a Escola particularmente interessada nos indicadores relativos a estas áreas.

A partir do pré levantamento de requisitos foram apresentados diversos *layouts* para serem discutidos e analisados em reunião.

Após esta fase foi realizada nova reunião para apresentação e posterior aprovação dos *Dashboards* finais.

### **3.4.1 REQUISITOS FUNCIONAIS E DE NEGÓCIO**

Os requisitos funcionais (RF) definidos para a aplicação desenvolvida baseiam-se nos indicadores fornecidos pela Direção da Escola e pelos responsáveis pelas áreas de Alunos e Docentes. Foram também auscultados de forma mais informal os coordenadores dos Diretores de Turma, os membros do Conselho Pedagógico e a Psicóloga da Escola. Além de consultas aos *stakeholders*, foram também obtidas informações a partir dos documentos estruturantes da Escola.

Todavia, numa primeira fase foi efetuada uma especificação de requisitos através da elaboração de uma análise funcional e posteriormente uma análise técnica. A análise funcional contempla todas as funcionalidades desejadas para este projeto:

- Informações das Escolas que compõem o Agrupamento;
- Informações de Alunos do Agrupamento;
- Informações de Docentes do Agrupamento;
- Informações de Funcionários do Agrupamento.

### **3.4.2 REQUISITOS NÃO FUNCIONAIS**

Os sistemas de *software* são cada vez mais complexos e os utilizadores cada vez mais exigentes. Hoje em dia não basta o sistema resolver funcionalidades exigidas, é necessário ir também ao encontro de aspetos funcionais pedidos pelos utilizadores tais como: custo, confiança, segurança, manutenção, portabilidade, performance, usabilidade entre outros. Todas estas características devem ser levadas em atenção desde o início do seu desenvolvimento e prolongando este tratamento por todo o ciclo de vida do *software* (Chung, Nixon, Yu, & Mylopoulos, 2000).

Requisitos não funcionais (RNF) passaram a ser exigidos pelos utilizadores para que um *software* seja considerado de qualidade. A nível do Agrupamento APJB os requisitos mais debatidos e considerados essenciais foram a precisão (precisão numérica), clareza da informação, interface amigável e o baixo custo.

### 3.5 MODELO DE DOMÍNIO

O Modelo de Domínio identifica os conceitos relacionados com requisitos do sistema e analisa o problema sob a perspectiva conceptual. Assim, o Modelo de Domínio deve ser independente da solução física que virá a ser adotada e deve conter apenas elementos referentes ao domínio do problema em questão, ficando para a fase de projeto os elementos da solução, isto é, todos os conceitos que se referem à implementação propriamente dita, tais como: interfaces, formas de armazenamento (base de dados), segurança de acesso, comunicação, entre outros.

Para a solução do projeto InfoPintor foi esboçado inicialmente o seguinte modelo:

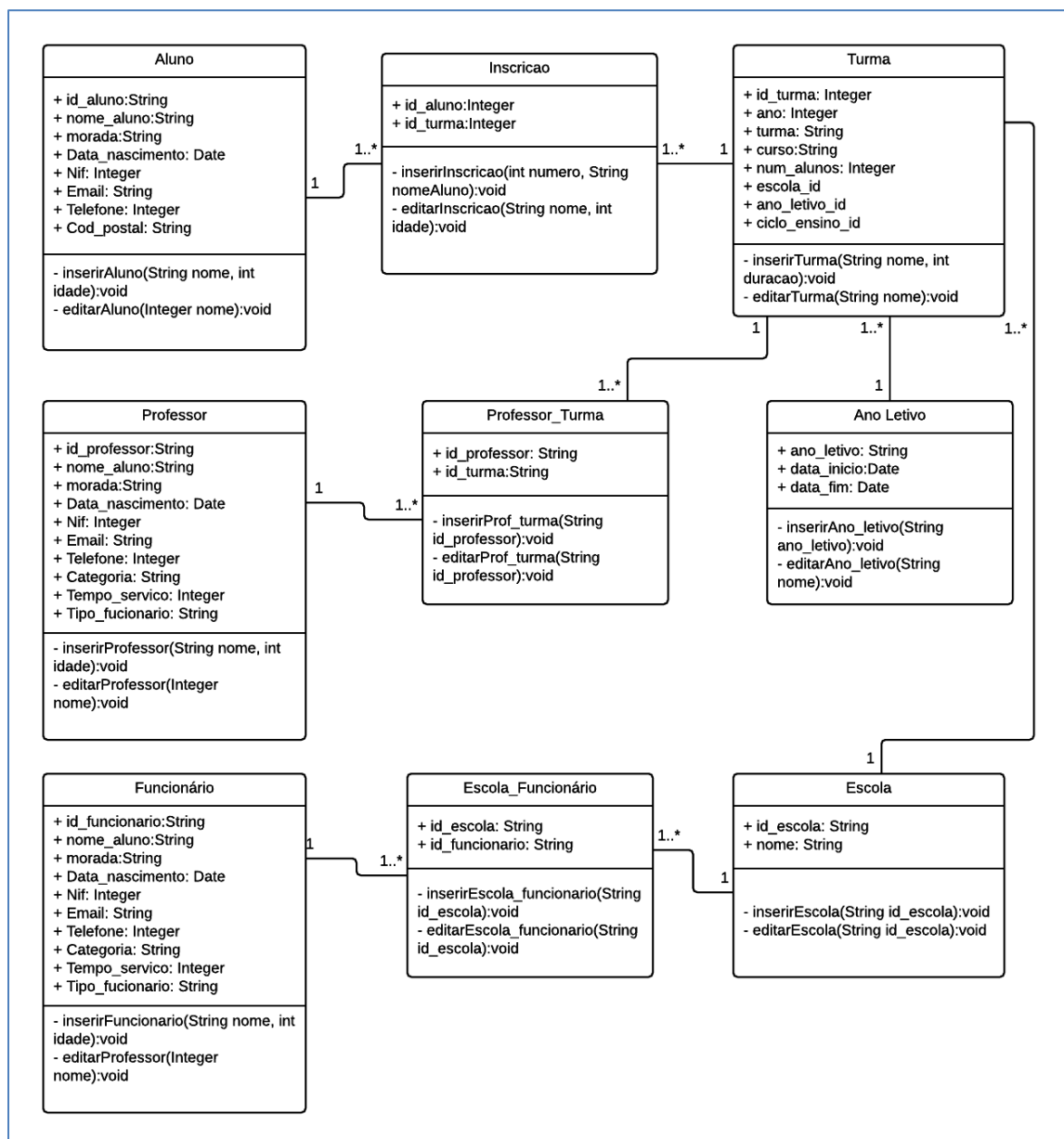


Figura 20: Modelo Domínio

A aplicação em análise, que foi implementada, trata de um caso real, descrevendo o modelo para gestão de uma Escola básica e secundária. O modelo apresenta quatro entidades principais: Alunos, Professores, Funcionários e Escolas. A nível de atributos, foram reduzidos ao mínimo indispensável para um funcionamento eficaz da gestão das informações.

### 3.6 CASOS DE USO

Uma vez identificados os requisitos, estes devem ser documentados para que possam sustentar o restante processo de desenvolvimento. Um diagrama de casos de uso mostra um conjunto de casos de uso, os atores intervenientes e os seus relacionamentos.

Um caso de uso descreve o comportamento do sistema sob certas condições, em resposta a uma requisição feita pelo *stakeholder* (Cockburn, 2005).

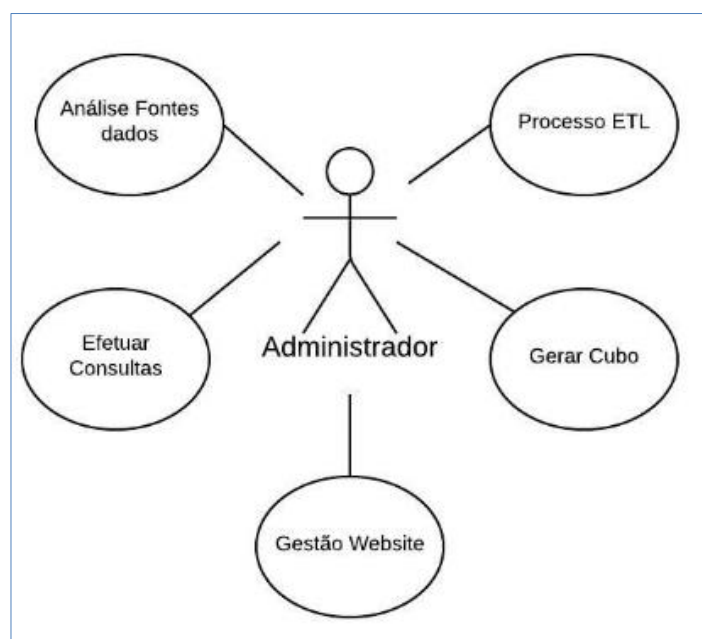


Figura 21: Diagrama Use Case (Admin)

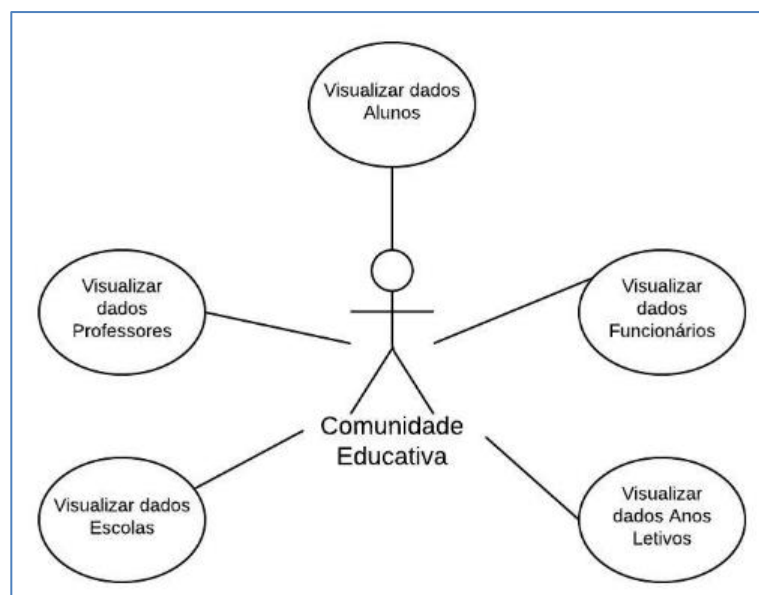


Figura 22: Diagrama Use Case (User)

No sistema em questão o DW tem dez casos de uso, existe um professor responsável (administrador) por analisar e extrair no início de cada ano letivo os dados das BD da Escola, executar o processo de ETL, gerar os cubos e efetuar a gestão do *website*. A Comunidade Educativa tem a possibilidade de efetuar consultas alargadas no *website*, bem como explorar os dados mais em detalhe na ferramenta Saiku.

### 3.7 ESPECIFICAÇÃO DA PROTOTIPAGEM

De acordo com os requisitos estipulados pela Escola, foram realizadas várias reuniões com a finalidade de cumprir os requisitos inicialmente estipulados, averiguar os dados a serem apresentados bem como a forma da sua apresentação. Para evitar constantes alterações ao longo do desenvolvimento do *software*, foram propostos protótipos idealizados para a aplicação. Esses protótipos passaram posteriormente por uma fase de validação, dissipando assim qualquer dúvida do que realmente era pretendido. A figura seguinte demonstra, como teste inicial, um exemplo de um protótipo apresentado aos órgãos de decisão da Escola no mês de janeiro que entretanto sofreu alterações pontuais durante o período de realização deste trabalho.



Figura 23: Protótipo da aplicação InfoPintor

### 3.8 CONSIDERAÇÕES FINAIS

Neste capítulo foi apresentado em linhas gerais o APJB. Foram também delineados os modelos de suporte à aplicação InfoPintor, que serão implementados no próximo capítulo com as necessárias adaptações à prática e às ferramentas utilizadas.

## 4. PLATAFORMA DE BI-DESENVOLVIMENTO

### 4.1 INTRODUÇÃO

O objetivo deste capítulo é apresentar o problema encontrado no Agrupamento APJB, expor a forma como será solucionado e descrever em detalhe os passos de desenvolvimento da solução proposta. (Jin & Fine, 1996) reconhecem que um interface com apresentação crucial de informação é uma mais-valia num ambiente de apoio à decisão. Incorporar tecnologias de visualização para o interface é uma abordagem promissora para apoiar a tomada de decisões. Por sua vez, independentemente da tecnologia usada, para os autores (Zhu & Chen, 2008), a visualização da informação é crucial para o sucesso de um sistema de apoio à decisão.

### 4.2 PROJETO TÉCNICO DA ARQUITETURA

Após o levantamento de requisitos, definidas as funções pretendidas e validação conjunta com a Direção do Agrupamento, foram estruturadas as seguintes funcionalidades:

- Repositório único (DW);
- Cubo analítico;
- Página Web *Dashboards* (Criação de *Dashboards* com os principais indicadores do Agrupamento que permitam uma visualização genérica das informações dos Alunos, Docentes e Funcionários);
- Indicadores (Indicadores de desempenho a calcular na análise OLAP após o levantamento de requisitos).

Depois de terem sido reunidos os requisitos da plataforma foi proposta a arquitetura do sistema devidamente adaptada ao contexto do projeto de BI a desenvolver na Escola. Na figura seguinte está representada a arquitetura técnica que permite a transformação dos dados de OLTP em *On-line Analytical Processing* (OLAP). A mesma figura ilustra a arquitetura do sistema segundo Kimball devidamente adaptada ao contexto da plataforma de BI implementada.

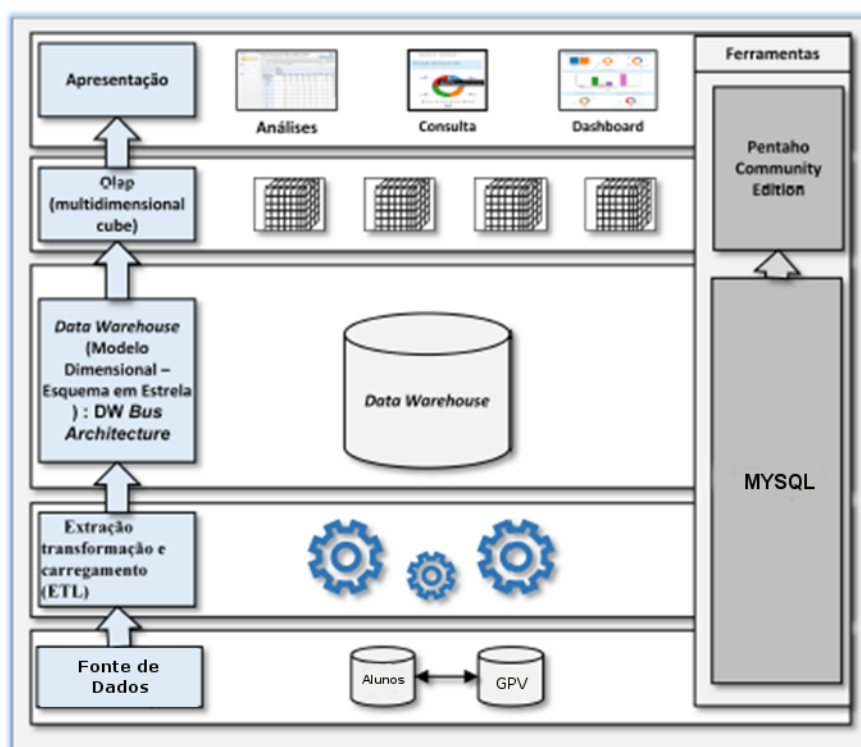


Figura 24: Arquitetura do sistema segundo Kimball

Fonte: Adaptado de (Kimball, 1998)

De uma forma geral e seguindo a metodologia de Kimball, os processos arquitetados e representados na figura anterior serão apresentados nas fases seguintes.

### 4.3 SELEÇÃO E ESCOLHA DAS FERRAMENTAS

As ferramentas de BI por norma são ferramentas caras, complexas e as empresas têm muita relutância na sua aposta. O Agrupamento não tinha disponibilidade financeira para tal aposta, pelo que, as ferramentas usadas neste projeto foram baseadas em modelos *open source*. Após um estudo comparativo das diversas plataformas (CF. Tabela 1), a plataforma BI escolhida foi a *Pentaho BI Suite*. O principal produto utilizado no projeto é o Pentaho<sup>2</sup>, que oferece a mais completa solução para BI.

A *Pentaho Corporation* comanda o papel de desenvolvimento do Projeto Pentaho BI - uma iniciativa da comunidade de desenvolvimento *Open Source* para oferecer

<sup>2</sup> *Open Source Business Intelligence*. Disponível em: <http://www.pentaho.com/>



ferramentas robustas de BI de código aberto. A suíte *Pentaho BI* é composta de ferramentas, de consultas e relatórios de gestão, análise de dados *OnLine Analytical Processing* (OLAP), tratamento de dados *Extraction, Transformation, Load* (ETL), mineração de dados (*Data Mining*) e *Workflow*.

O *software* Pentaho mostrou-se adequado e respondeu a todas as questões da manipulação do sistema InfoPintor, destacando a facilidade de construção e de acesso a todas as funcionalidades e resultados. A ferramenta faz uma gestão dinâmica por meio de uma interface web amigável e intuitiva. Esta ferramenta permite também carregar dados de diferentes fontes (xls, csv, xml, entre outros).

No contexto de ETL, foi definido como objetivo estratégico o uso de uma solução livre de BI para fazer face às solicitações desejadas. Após análise de diversa documentação, das características de algumas ferramentas, das suas mais-valias e das suas fragilidades, a escolha da ferramenta para este projeto recaiu sobre o Pentaho Data Integration (PDI), também conhecida como *Kettle*. O DM produzido foi armazenado no SGBD livre MySQL.



Figura 25: Tecnologias Usadas no Desenvolvimento da Aplicação InfoPintor

Para tratamento de consultas OLAP, a geração de cubos no Pentaho, foi feito o mapeamento de um ficheiro em XML, facilmente interpretado pelo servidor OLAP. A ferramenta usada para a análise dos Cubos OLAP gerados, foi o *Analysis View* (Mondrian) que permite a construção de cubos que representam modelos de dados multidimensionais. Com a adição do *plugin* CDE, foi possível definir o *layout* da aplicação web, como por exemplo, adicionar caixas de seleção, caixas de texto, tabelas, gráficos, entre outros.

Na aplicação web InfoPintor são disponibilizados indicadores de desempenho através de *Dashboards*. Para apresentação do *layout* gráfico foi utilizado o CDA que faz uso de ferramentas típicas da web (HTML e Javascript) e que permite uma grande flexibilidade.

#### 4.3.1 INSTALAÇÃO DO PENTAHO

A instalação da solução Pentaho precisa de diversas configurações mediante as ferramentas necessárias para a aplicação a desenvolver.

Após *download* é necessário proceder à configuração da máquina Java para o Pentaho. Posteriormente é necessário proceder à configuração da ligação à BD. Finalizado este processo o Pentaho encontra-se preparado para ser utilizado. A versão usada nesta solução InfoPintor é a versão gratuita, assim, foi necessário descarregar a ferramenta *Data Integration* (ferramenta que permite o processo ETL), instalar o *plugin* Saiku (testar o cubo) e o *plugin* Dashboard CDE (criação de painéis) via Marketplace.

#### 4.4 DESENVOLVIMENTO DO SISTEMA ETL E DA ESTRUTURA DA PLATAFORMA DE BI

O mecanismo de integração e consulta InfoPintor é descrito em três fases (CF. Figura 25). A Fase 1 descreve os possíveis dados a integrar na aplicação, a fase 2 foca o processo de ETL e a criação dos DM e por último, na fase 3, são demonstradas as técnicas de análise e exploração dos dados InfoPintor.

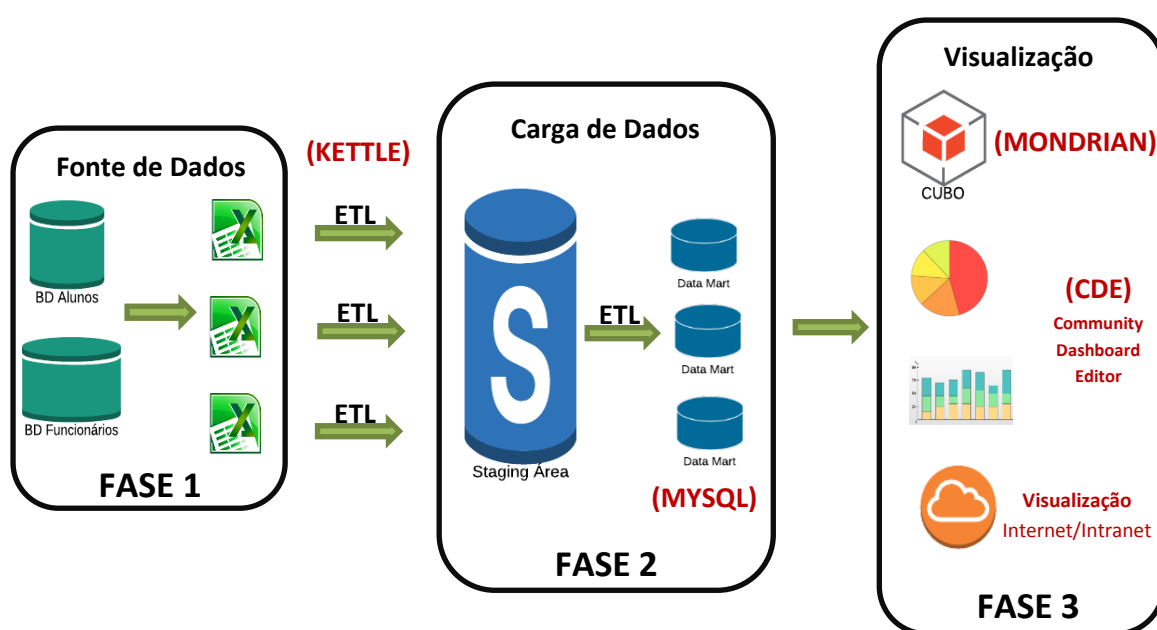


Figura 26: Esquema do Processo de Desenvolvimento da Aplicação

No processo de implementação foram seguidos os seguintes passos:

- Levantamento das fontes de dados: Em primeiro lugar, foi feito o estudo das fontes de dados necessárias para o sistema InfoPintor e um levantamento dos sistemas existentes, pois, só após este estudo, foi possível definir o caminho a seguir. A primeira fase desse processo é a extração dos dados das bases de dados ALUNOS (proveniente do programa JPM Alunos) e a base de dados DOCENTES/FUNCIONÁRIOS (proveniente do programa GPV – Gestão de Pessoal e Vencimentos). A base de dados ALUNOS, constituída por 5365 registos, é composta basicamente por metadados referentes aos dados pessoais do aluno bem como ao seu percurso escolar. Os dados dos alunos são agrupados por ciclos com identificação do número do processo do aluno, disciplina, ano letivo de conclusão do ciclo e a respetiva avaliação. A base de dados DOCENTES/FUNCIONÁRIOS, constituída por 1223 registos, é composta na sua maioria por metadados referentes aos dados dos professores e funcionários, bem como à sua situação profissional. Para a implementação deste projeto no contexto educativo do Agrupamento, foi analisado um conjunto de situações em que pudessem estar em causa o direito, salvaguarda e proteção dos dados dos alunos, professores e funcionários na ótica legal, ética, social, pedagógica e profissional.
- Exportar os dados para formato *Excel*: Os dados das BD Alunos e Docentes/Funcionários foram exportados para ficheiros de *Excel* por forma a serem melhor interpretados pelo processo ETL. Usando a ferramenta *Kettle*, procedeu-se à extração da informação de forma integral para um *staging area* em MYSQL. Aquando a exportação é feito o anonimato dos dados, sendo apenas registado o número de processo do aluno, do docente e do funcionário, desaparecendo qualquer tipo de referência ao nome. A partir deste momento o aluno apenas é conhecido pelo seu número de processo.

Numa segunda fase, (Kimball, 1998) no seu livro *The Data Warehouse Toolkit*, refere a importância de seguir um modelo de escolhas consecutivas para projetos com estruturas dimensionais:

- 1-Escolha do processo de negócio;
- 2-Escolha do nível de granularidade;
- 3-Seleção das dimensões;
- 4-Seleção dos factos.

Para a projeção do DW foi usada a modelação dimensional da BD, cujo modelo em estrela é o que mais se adequa à estrutura do DW pretendido. Este modelo é composto por dois tipos de tabelas: Factos e Dimensões.

No esquema em estrela, os factos são representados na tabela de factos que ocupa a posição central do diagrama. Por sua vez, as dimensões são representadas nas diversas tabelas de dimensão, estas tabelas circundam a tabela de factos. As dimensões usadas na construção do DW InfoPintor estão abaixo especificadas.

<b>Dimensão</b>	<b>Descrição</b>
Dim_escola	A dimensão escola representa todas as escolas do Agrupamento APJB.
Dim_ano_letivo	Esta dimensão é crucial em qualquer DM. É utilizada para limitar intervalos de tempo.
Dim_subsidio	Dimensão que regista se um aluno é ou não subsidiado e qual o escalão que lhe foi atribuído.
Dim_sexo	Dimensão que regista a opção M (Masculino) e F (Feminino) para funcionários, alunos e docentes.
Dim_ciclo	A dimensão ciclo representa os 5 ciclos de ensino existentes: 0 - Pré Escolar 1 - 1ºciclo 2 - 2ºciclo 3 - 3ºciclo 4 - Secundário 5 – Profissional e Vocacional
Dim_funcionario	A dimensão funcionário representa os recursos humanos da escola (Funcionários e Docentes).
Dim_faixa_etaria	Esta dimensão representa as várias faixas etárias em intervalos de 5 anos.
Dim_aluno	Esta dimensão guarda informações dos diversos alunos, como a data de nascimento, telefone, nacionalidade, entre outros.
Dim_regime	Dimensão que informa se o funcionário pertence ao grupo dos docentes ou dos não docentes.
Dim_grupo	Dimensão que apresenta o grupo de recrutamento dos docentes.
Dim_categoria	Esta dimensão representa a categoria profissional dos funcionários do agrupamento.

Dim_tempo_serviço	Dimensão que representa o tempo de serviço dos docentes e dos funcionários.
Dim_turma	Dimensão que representa a turma do aluno.

**Tabela 2: Tabelas de dimensão do DW**

A chave primária da tabela de factos é composta pelas diversas chaves estrangeiras das tabelas de dimensão, sendo que os únicos relacionamentos possíveis são entre a tabela de facto e as dimensões, não sendo possível ligações diretas entre tabelas de dimensão. As tabelas de facto usadas na construção do DW InfoPintor estão abaixo especificadas.

<b>Facto</b>	<b>Descrição</b>
F_ciclo_ensino	Factual que guarda informações sobre o Agrupamento.
F_funcionarios	Factual que guarda informações sobre os funcionários.
F_inscritos	Factual que guarda informações sobre os alunos.

**Tabela 3: Tabelas de factos do DW**

Em termos de medidas numéricas (*measures*) são definidas em função dos valores que se deseja mostrar na ferramenta OLAP. Para a aplicação InfoPintor foram definidas as seguintes métricas:

<b>Métricas</b>	<b>Descrição</b>
Count	Contagem por exemplo do número de alunos
Med_idades	Média por exemplo das idades dos alunos
Max	Máximo tempo de serviço

**Tabela 4: Tabela de métricas**

Para o projeto InfoPintor foi projetado o seguinte DW. O modelo de dados proposto é composto por 3 tabelas de factos e 14 tabelas de dimensão. Os assuntos que compõem as tabelas de factos foram devidamente categorizados para irem ao encontro do que a Escola pretende. Somente após esta base estar construída se procedeu à elaboração do ETL.

Em termos de processo ETL, foi objetivo primordial neste projeto definir os esquemas de ETL de forma simples, recorrendo a componentes pouco complexos, de forma a serem facilmente interpretados e de execução rápida. Após análise dos dados recolhidos das duas bases de dados da Escola foram projetados três DM.

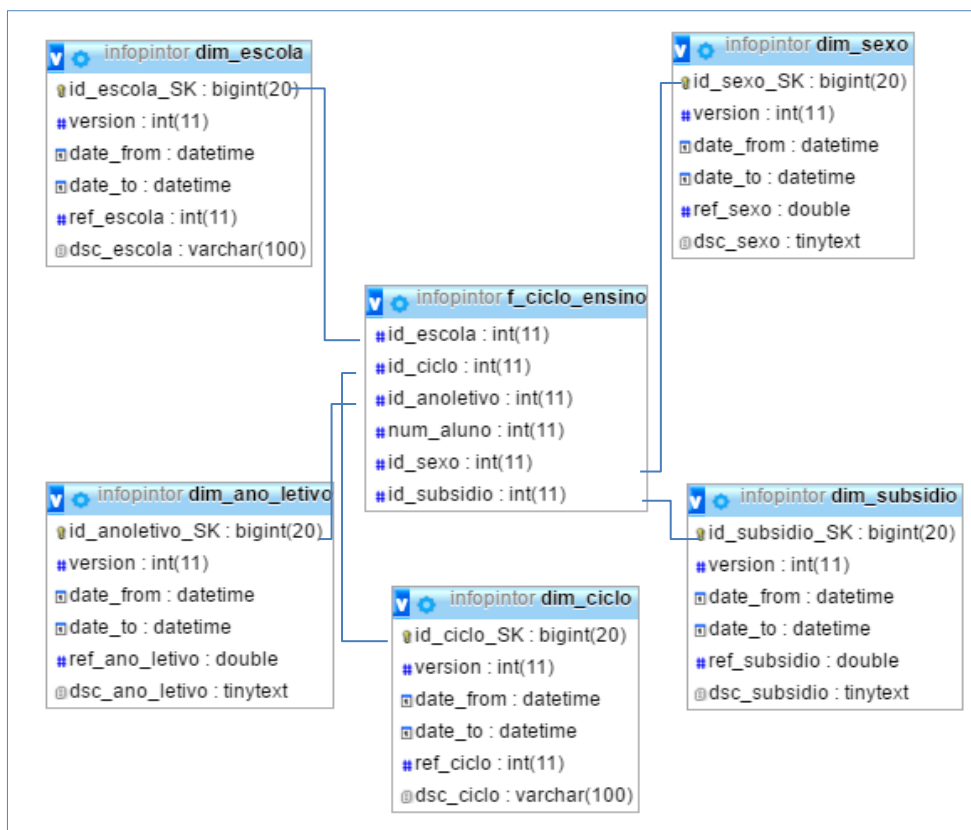


Figura 27: DM Ciclo Ensino

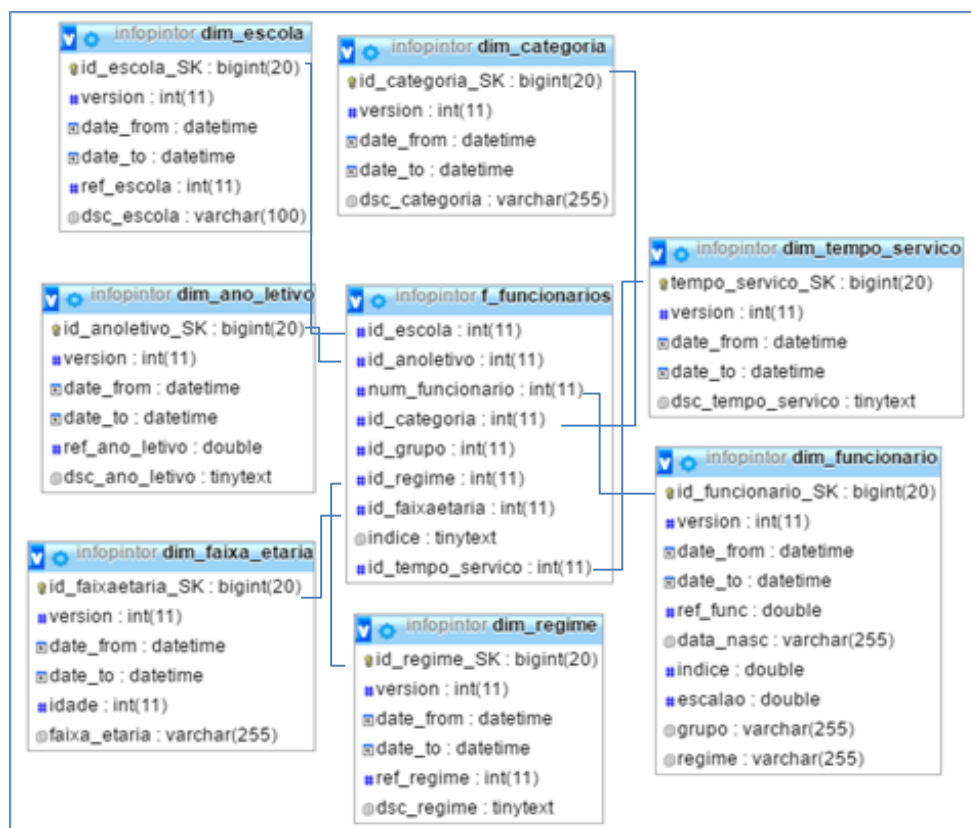


Figura 28: DM Funcionário

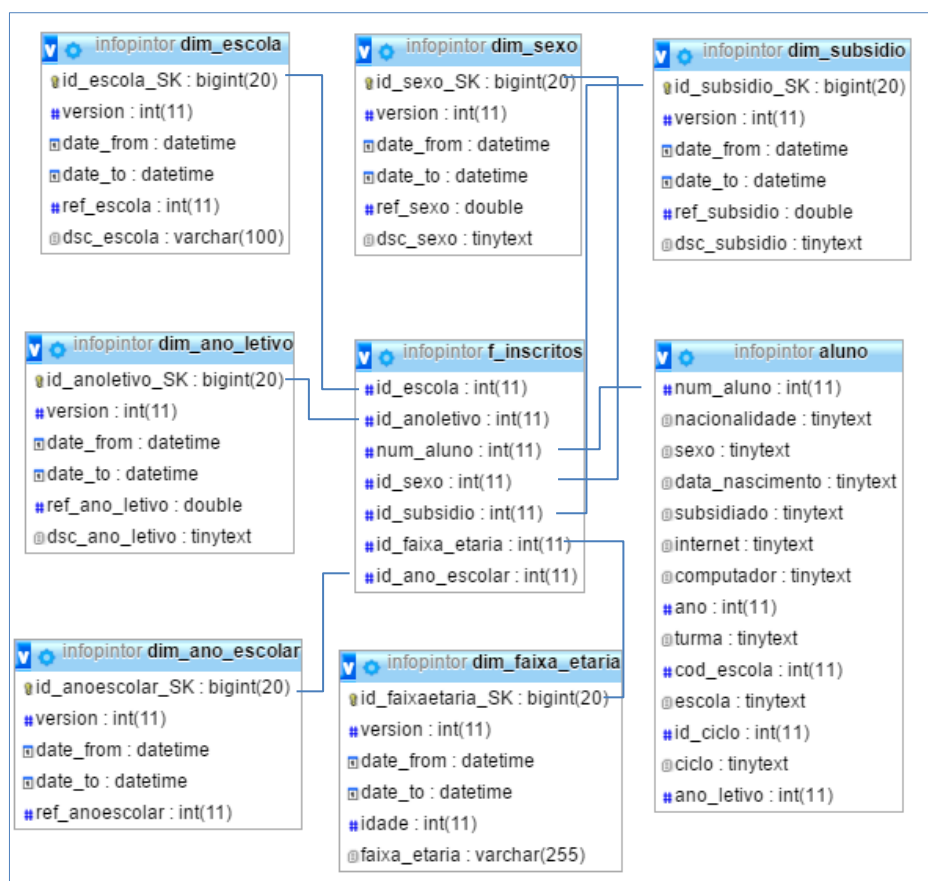


Figura 29: DM Inscritos

A ferramenta escolhida da *suite* Pentaho foi o *Pentaho Data Integration* (PDI), originalmente denominada de *Kettle*. Esta ferramenta é responsável por realizar o processo de ETL de diversos tipos de dados. As transformações ETLs (uma transformação é uma rotina com uma coleção de passos interligados, dos quais o primeiro representa a fonte de dados e o último representa a saída dos dados) foram divididas em quatro grupos: ETLs das tabelas de *staging*; ETLs das tabelas dimensão; ETLs das tabelas de facto e por último ETLs de otimização de processos.

- Carregamento na área temporária (*Staging Area*): A partir dos ficheiros em *Excel* e através de um processo ETL simples foi feito o carregamento dos dados na *staging area*. As tabelas resultantes contêm informação temporária e posteriormente auxiliam no preenchimento das dimensões e dos factos. A seguinte transformação retrata a criação da tabela de *staging* Aluno. Nesta transformação os campos do ficheiro em *Excel* são analisados, e com o componente *Script Java* é convertida a data de nascimento para o formato AA/MM/DD e o resultado final é carregado numa tabela denominada Aluno na BD InfoPintor.

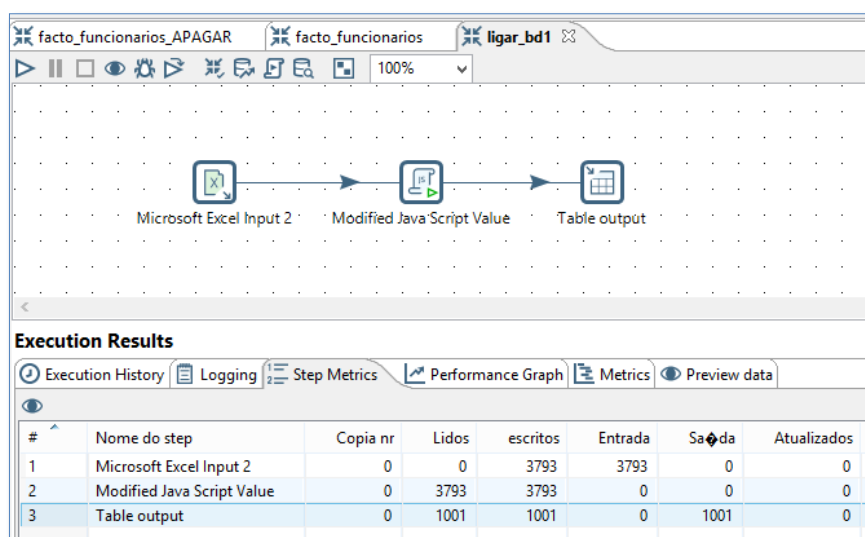


Figura 30: Esquema de Transformação Alunos

A seguinte transformação retrata a criação da tabela de staging Funcionário. A partir de dois ficheiros em Excel (Funcionários e Docentes) são analisados e seleccionados os campos pretendidos, é convertida a data de nascimento para o formato AA/MM/DD e o resultado final é carregado numa tabela denominada Funcionário na BD InfoPintor.

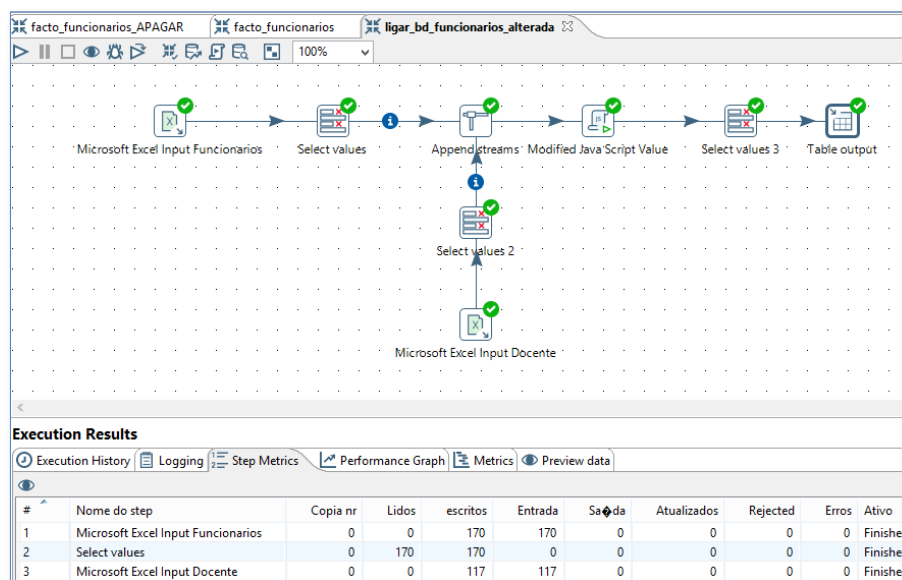


Figura 31: Esquema de Transformação Funcionários

- Carregamento das dimensões: Por uma questão de organização foram mantidas na base de dados as tabelas de *staging* e as novas tabelas de dimensão e de factos, usando terminologias diferentes para os nomes das tabelas do DW.



Dimensão Aluno - A partir da tabela de *staging* Aluno é verificado se existem campos nulos e são selecionados os campos pretendidos. O componente *Dimension lookup/update* é deveras importante neste processo pois guarda um conjunto de dados numa tabela, cria a chave primária automaticamente e permite o controlo de versões. É utilizado maioritariamente no armazenamento de dados das dimensões.

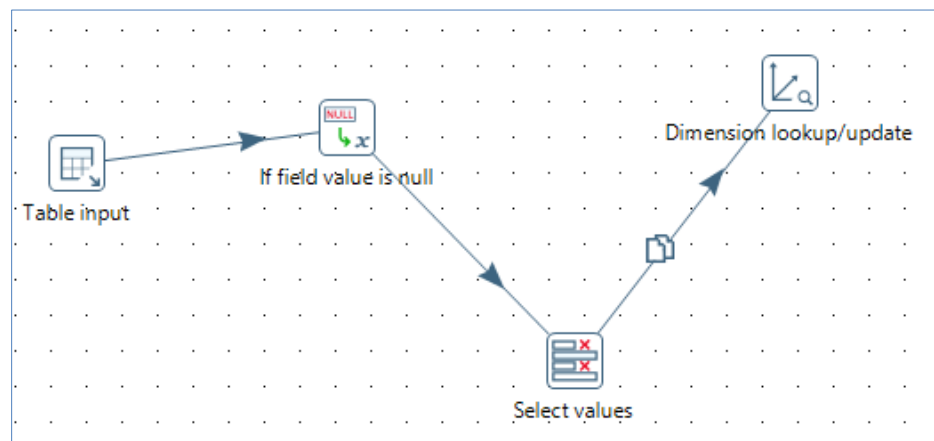


Figura 32: Criar Dimensão Aluno (Dim\_Aluno)

Dimensão Ano Escolar - O componente *Generate Rows* permite criar registos, neste caso 13 linhas correspondentes aos 13 anos escolares existentes no ensino, desde 0 (Pré Escolar) até 12 (12º ano). Com o componente *Add sequence* é gerada a sequência dos anos escolares e por fim é feita a gravação dos dados na tabela de dimensão correspondente.

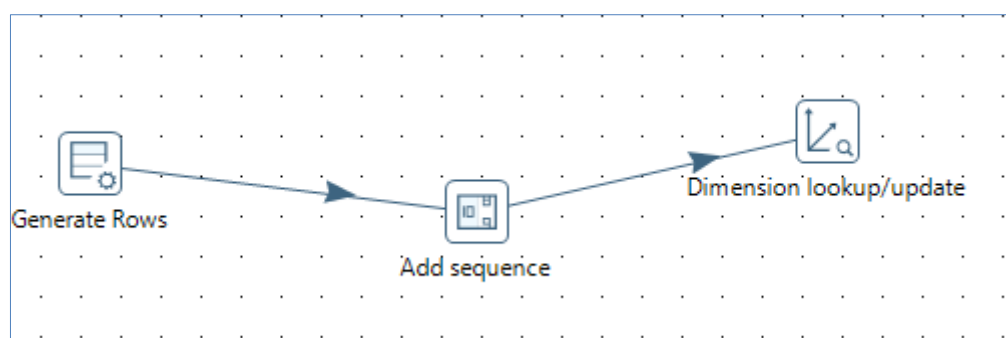


Figura 33: Criar Dimensão Ano Escolar (Dim\_Ano\_Escolar)

Dimensão Categoria - Transformação simples que a partir de um ficheiro em *Excel* o analisa, e selecciona os valores pretendidos para gravação na tabela de dimensão.

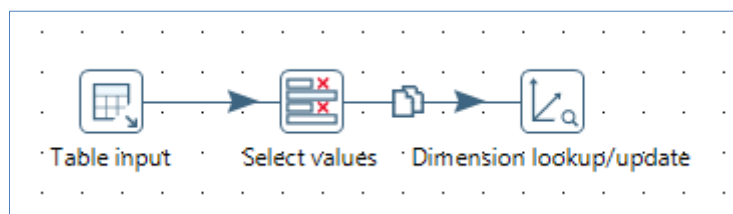


Figura 34: Criar Dimensão Categoria

Dimensão Tempo de Serviço - A partir do ficheiro em *Excel* que contém a data de início de funções e a data de cessação do contrato de um funcionário, foi utilizado o componente *Script Java* para calcular o tempo de serviço. Por fim foram seleccionados os campos pretendidos e feita a gravação na respetiva dimensão.

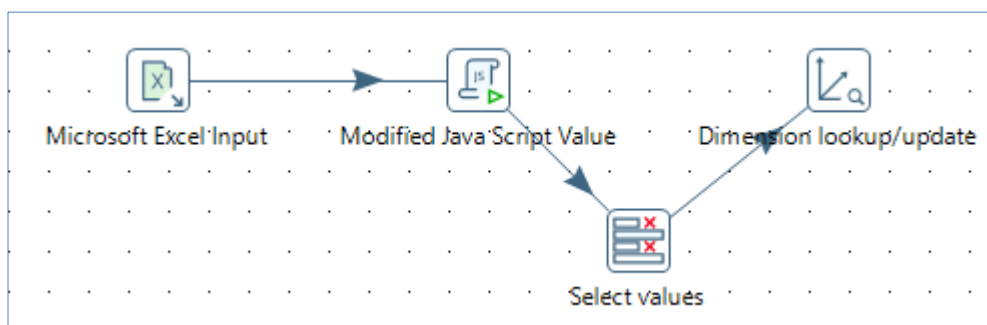


Figura 35: Criar Dimensão Tempo Serviço

Dimensão Faixa Etária - O componente *Generate Rows* permite criar registos, neste caso foram criadas 120 linhas. Com o componente *Add sequence* é gerada a sequência dos anos (de 1 a 120) e o componente *Number range* permite criar os intervalos para a faixa etária. Por fim é feita a gravação dos dados na tabela de dimensão correspondente.

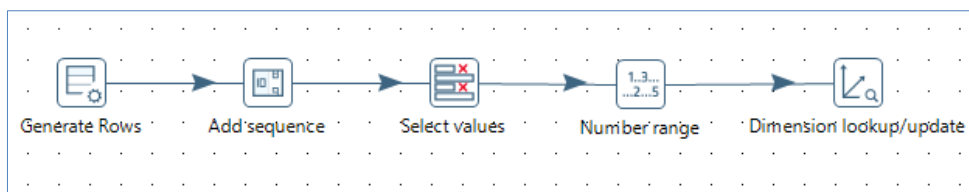
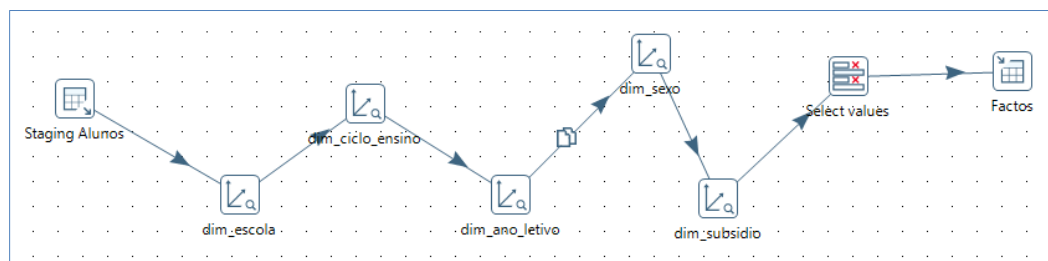


Figura 36: Criar Dimensão Faixa Etária

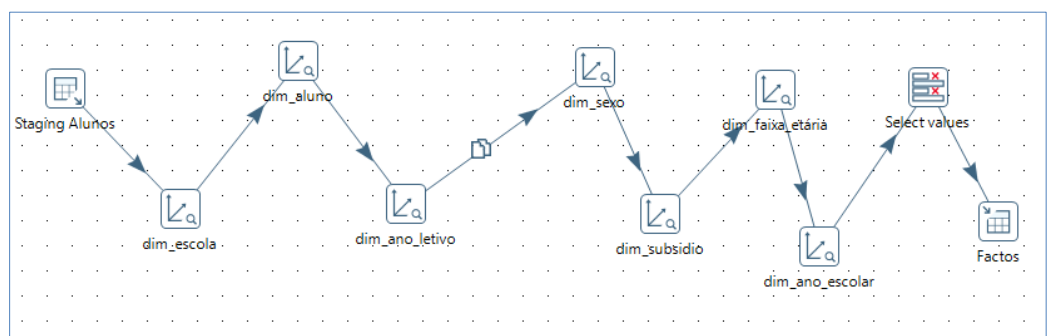
▪ Carregamento dos factos: Este passo é considerado a parte final do carregamento do DW. No modelo dimensional acima apresentado, foram adicionadas colunas responsáveis por fazer a ligação entre a tabela de facto e as suas dimensões. Foi usado o prefixo "\_SK" no modelo dimensional, como por exemplo "id\_ciclo\_SK".

Factual Ciclo de Ensino - A partir da tabela de staging Alunos são adicionadas as chaves provenientes das várias dimensões e por fim criada a tabela factual de "f\_ciclo\_ensino". A tabela "f\_ciclo\_ensino" contém dois tipos de dados, as métricas e as chaves estrangeiras que permitem a ligação a cada uma das tabelas de dimensão.



**Figura 37: Criar Factual Ciclo Ensino**

Factual Inscritos - A partir da tabela de *staging* Alunos são adicionadas as chaves provenientes das várias dimensões e criada a tabela factual de "f\_inscritos". Esta factual armazena os dados respetivos aos alunos por ano escolar.



**Figura 38: Criar Factual Inscritos**

Factual Funcionários - A partir da tabela de *staging* Funcionários são selecionados apenas os registos pretendidos para o sistema em questão, são

adicionadas as chaves provenientes das várias dimensões e por fim criada a tabela factual “f\_funcionarios”.

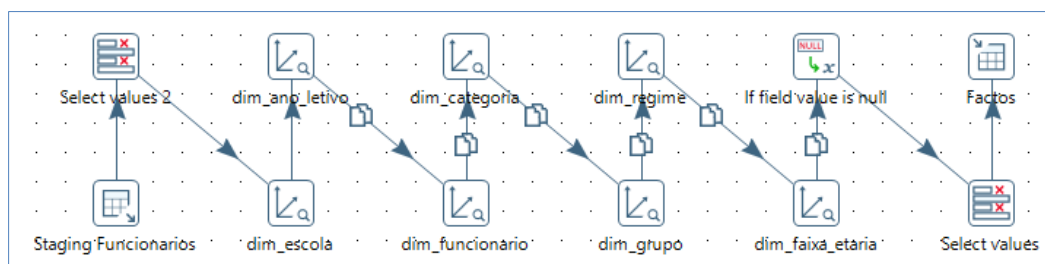


Figura 39: Criar Factual Funcionário

- Otimização do ETL: Foi criado um novo *Job* (*Job* permite correr dimensões e factuais entre outras funcionalidades) para executar, de uma só vez, todas as dimensões existentes no sistema a implementar.

*Job* dimensões – *Job* que executa automaticamente todas as dimensões não sendo necessário executá-las uma a uma.

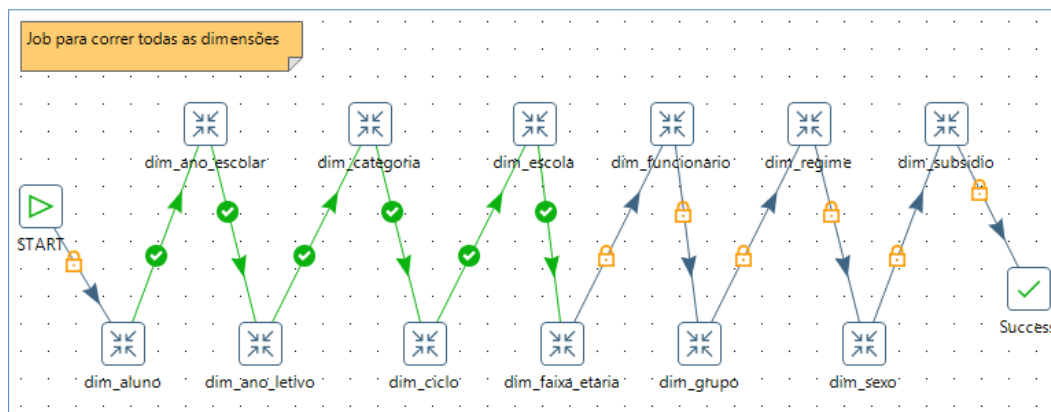


Figura 40: Job Dimensões

*Job* factuais – *Job* que executa automaticamente todas as factuais não sendo necessário executá-las uma a uma.

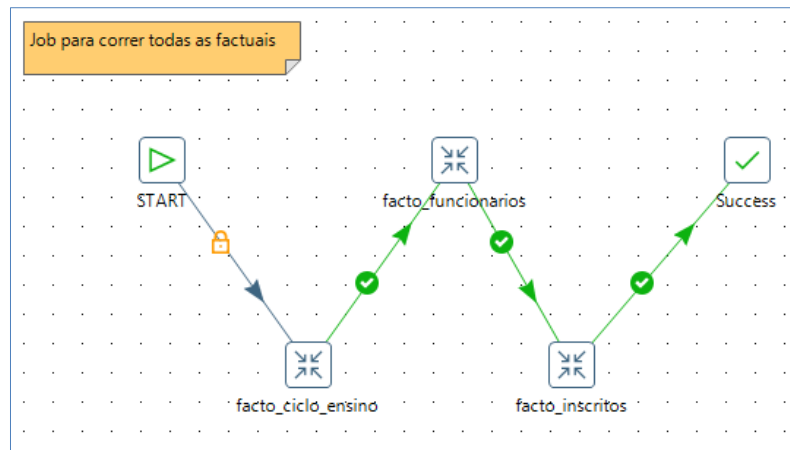


Figura 41: Job Factuais

*Refresh* ao Mondrian – Transformação que efetua *refresh* à *cache* do Mondrian. Esta operação é importante pois a cada novo carregamento no DW torna-se necessário atualizar os dados no Mondrian. Foi escolhida uma solução genérica, ou seja, o administrador efetua *login* no BI Server e poderá atualizar os dados.

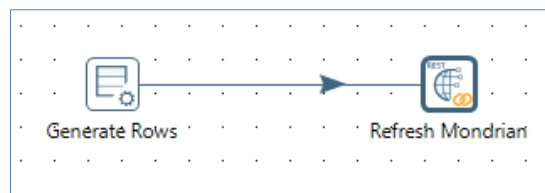


Figura 42: Transformação Refresh Mondrian

*Job Geral* para correr todas as dimensões, factuais e *refresh* ao Mondrian.

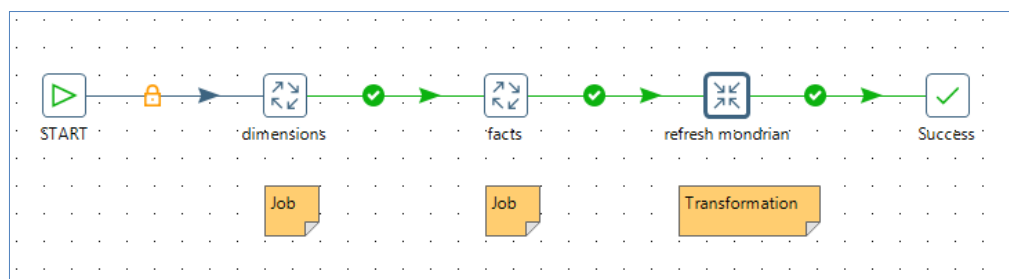


Figura 43: Jog Geral

Executar script: Foi criado um ficheiro executável (run\_chain.bat) que permite executar o *job* anterior (job geral).

```
@echo off  
  
cd ../../../../Pentaho/pdi-ce-5.4.0.1-130/data-integration  
  
kitchen.bat /file="../../../../PROJETO FINAL/8-Processo ETL (DW)/ETL Tratados/chain_principal.kjb"
```

Figura 44: Script Genérico

Após o processo ETL já é possível visualizar a BD com os seus principais componentes (tabelas de factos, tabelas de dimensão).

Nesta etapa é importante garantir a segurança e a fiabilidade dos dados para tomadas de decisão ágeis e eficazes. Os dados são carregados na BD do DW segundo um modelo multidimensional.

A fase 2 termina assim com a implementação de três DM (Ciclo Ensino, Alunos e Funcionários) em MYSQL.

Em termos da terceira fase:

- Cubos OLAP: Depois de armazenados nos DM, os dados podem ser manipulados por ferramentas de BI, podem ser visualizados em cubos OLAP, em relatórios ou em *Dashboards*. Estas ferramentas analisam e apresentam a informação de uma forma direta e objetiva aos utilizadores finais da aplicação.

Um dos segmentos mais populares dos aplicativos de BI é o que lida com OLAP, um termo muito usado para denominar aplicações que possibilitam uma visão multidimensional de dados. O paradigma de consulta aos dados é diferente do modelo convencional baseado num modelo puramente relacional e nas aplicações transacionais convencionais (Pessoa, 2002).

Numa primeira etapa foi necessário configurar o cubo OLAP no ficheiro de configuração XML, normalmente apelidado de *Mondrian Cube Schema*. Neste ficheiro são definidas as dimensões (simples e partilhadas), as factuais e os níveis de hierarquia das dimensões. Existem aplicações *open source* como o *Pentaho Cube Designer* e o *Schema Workbench* que auxiliam no processo de criação dos cubos, no entanto, o tempo disponível para a concretização do projeto não foi suficiente para explorar mais uma ferramenta. Para a definição do esquema optou-se, neste projeto, por não recorrer a nenhuma interface gráfica, utilizando o *Notepad++* para criação do esquema Mondrian em XML.

O esquema Mondrian é o ficheiro que diz ao OLAP *engine* (mecanismo OLAP) Mondrian (nome do motor) como interpretar a BD e exibir um cubo OLAP. O modelo da base de dados não é suficiente para o Pentaho identificar os cubos e as dimensões. É necessário criar um arquivo XML descrevendo quais são os cubos, dimensões e as respetivas tabelas associadas. Somente após a criação deste ficheiro XML é possível gerar uma informação de forma a ser identificado pelo Pentaho.

Como referido anteriormente, cada um destes cubos representa um modelo “em estrela” e são constituídos por uma tabela de factos e pelas respetivas dimensões que a caracterizam. A estrutura genérica para a construção de um cubo é a demonstrada na seguinte figura:

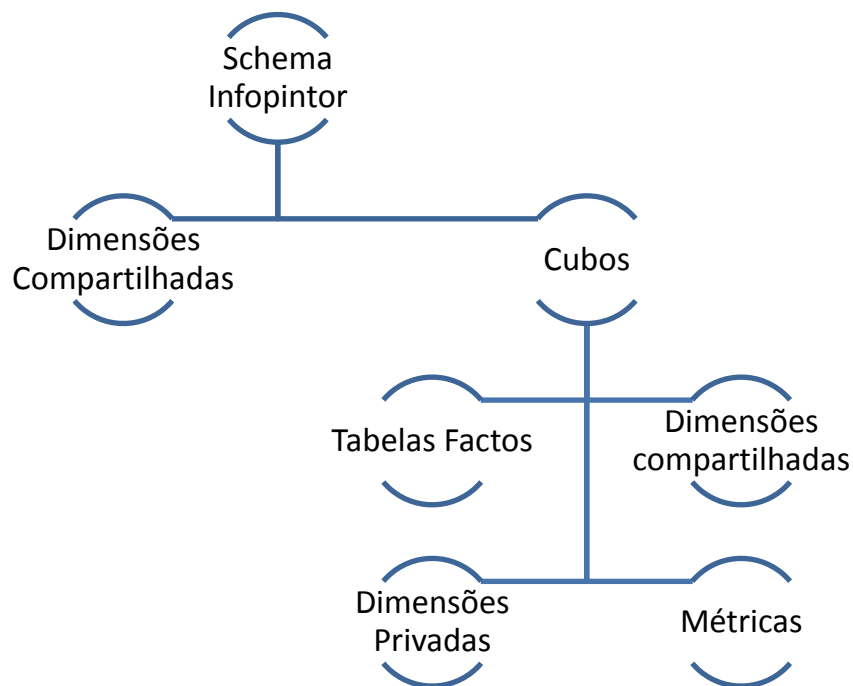


Figura 45: Estrutura Construção de um Cubo OLAP

Inicialmente deve ser criado um *schema* Mondrian tendo por base a BD criada anteriormente. Seguem-se as dimensões compartilhadas que são dimensões que estão disponíveis para qualquer cubo na BD OLAP e a construção do cubo propriamente dito.

Cada cubo deve ser baseado somente numa tabela de factos. As métricas representam as colunas provenientes da tabela de factos que contém os dados numéricos que se pretendem analisar. É possível adicionar ao cubo o número de medidas que se pretenderem, desde que, essas colunas estejam presentes na tabela

de factos. É possível também criar o número de cubos que se pretender, cada cubo pode ter o seu próprio conjunto de dimensões e medidas.

A estrutura do *schema* deste projeto foi pensada para ir ao encontro do pretendido pela Escola como pode ser observado abaixo no extrato do ficheiro em XML. Unicamente a partir deste XML é que se pode gerar uma informação do tipo XACTION, que é o arquivo que será identificado pelo Pentaho. Para OLAP optou-se pela ferramenta Notepad++ que permite a criação de cubos, dimensões e medidas de uma forma bastante simples.

```
<Schema name="INFOPINTOR">
  <!-- Dimensoes partilhadas -->
    <Dimension caption="Escola" highCardinality="false"
name="dim_escola_shared" type="StandardDimension" visible="true">
      <Hierarchy allMemberCaption="Todos" caption="Escola"
allMemberName="Todos" hasAll="true" name="dim_escola"
primaryKey="id_escola_SK" visible="true">
        <Table name="dim_escola"></Table>
        <Level caption="Escola" name="nome_escola"
hideMemberIf="Never" internalType="int" levelType="Regular"
column="ref_escola" nameColumn="dsc_escola" ordinalColumn="ref_escola"
type="Integer" uniqueMembers="false" visible="true">
          </Level>
        </Hierarchy>
      </Dimension>
    <Dimension caption="AnoLetivo" highCardinality="false"
name="dim_ano_letivo_shared" type="StandardDimension" visible="true">
      <Hierarchy allMemberCaption="Todos" caption="AnoLetivo"
allMemberName="Todos" hasAll="true" name="dim_ano_letivo"
primaryKey="id_anoletivo_SK" visible="true">
        <Table name="dim_ano_letivo"></Table>
        <Level caption="AnoLetivo" column="ref_ano_letivo"
nameColumn="dsc_ano_letivo" hideMemberIf="Never" internalType="int"
levelType="Regular" name="anoletivo" type="Integer"
uniqueMembers="false" visible="true">
          </Level>
        </Hierarchy>
      </Dimension>
    (...)
  <!-- CUBO CICLO ENSINO -->
    <Cube cache="true" caption="CicloEnsino_Cube" enabled="true"
name="cicloensino" visible="true">
      <Table name="f_ciclo_ensino"></Table>
      <!-- Factual -->
      <DimensionUsage name='dim_escola' source="dim_escola_shared"
foreignKey="id_escola"></DimensionUsage>
```



```

        <DimensionUsage name='dim_ano_letivo'
source="dim_ano_letivo_shared"
foreignKey="id_anoletivo"></DimensionUsage>
        (...)
        <Dimension caption="CicloEnsino" highCardinality="false"
foreignKey="id_ciclo" name="dim_ciclo_shared" type="StandardDimension"
visible="true">
        <Hierarchy allMemberCaption="Todos" caption="CicloEnsino"
allMemberName="Todos" hasAll="true" name="dim_ciclo"
primaryKey="id_ciclo_SK" visible="true">
        <Table name="dim_ciclo"></Table>
        <Level caption="CicloEnsino" column="ref_ciclo"
hideMemberIf="Never" internalType="int" levelType="Regular"
name="nome_ciclo" nameColumn="dsc_ciclo" ordinalColumn="ref_ciclo"
type="Integer" uniqueMembers="false" visible="true">
        </Level>
        </Hierarchy>
        </Dimension>
        <Measure aggregator="distinct-count" caption="Numero de
Alunos" column="num_aluno" name="numero_alunos"
visible="true"></Measure>
    </Cube>
    (...)
</Schema>

```

Para testar o cubo acima apresentado, recorreu-se à ferramenta Pentaho BI Server (Saiku). Foi escolhido o Saiku após pesquisas que o apontavam como o melhor visualizador OLAP a funcionar com o Mondrian/Pentaho. O Saiku *Analytics* é um cliente web disponível como *plugin* para o Pentaho BI Server. Usa o Pentaho Analysis (Mondrian) para proporcionar uma forma fácil e amigável de aceder e explorar os recursos OLAP dos cubos multidimensionais, facilitando assim a vida ao utilizador final.

Exploração visual do cubo com Saiku: Foi instalado o *plugin* Saiku via Marketplace para proporcionar uma visualização dos dados do DW/Cubo de uma forma mais dinâmica e amigável.

Community Dashboard Editor Webdetails	2	Installed 5.4.0.1-130 (5.4.0.1)	Up to Date	✓
Community Graphics Generator Webdetails	2	Installed 5.4.0.1-130 (5.4.0.1)	Up to Date	✓
Sparkl - Pentaho Application Builder Webdetails	3	Installed master (5.4.0.1)	Up to Date	✓
Saiku Analytics Meteorite BI	4	Installed 3.7 (TRUNK)	Up to Date	✓

Figura 46: Instalação do Saiku

O Saiku não é mais que uma ferramenta analítica, baseada em tecnologia web, que permite criar relatórios e explorar a informação contida nos cubos. A grande mais-valia desta ferramenta é permitir a possibilidade de explorar e extrair informação por parte da Comunidade Educativa em geral sem ser necessário técnicos especializados ou com conhecimentos aprofundados da ferramenta. Neste caso específico, é possível explorar os três cubos anteriormente criados.

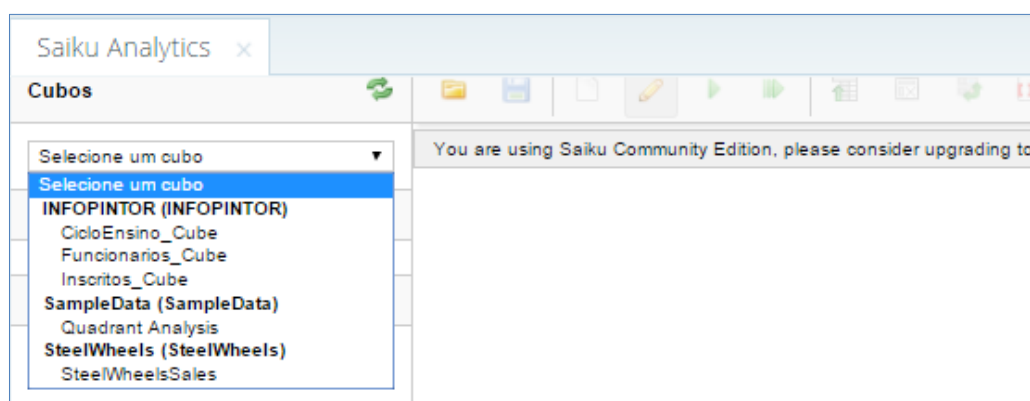


Figura 47: Exploração no Saiku

Na imagem seguinte é possível visualizar a exploração do cubo funcionários no Saiku.

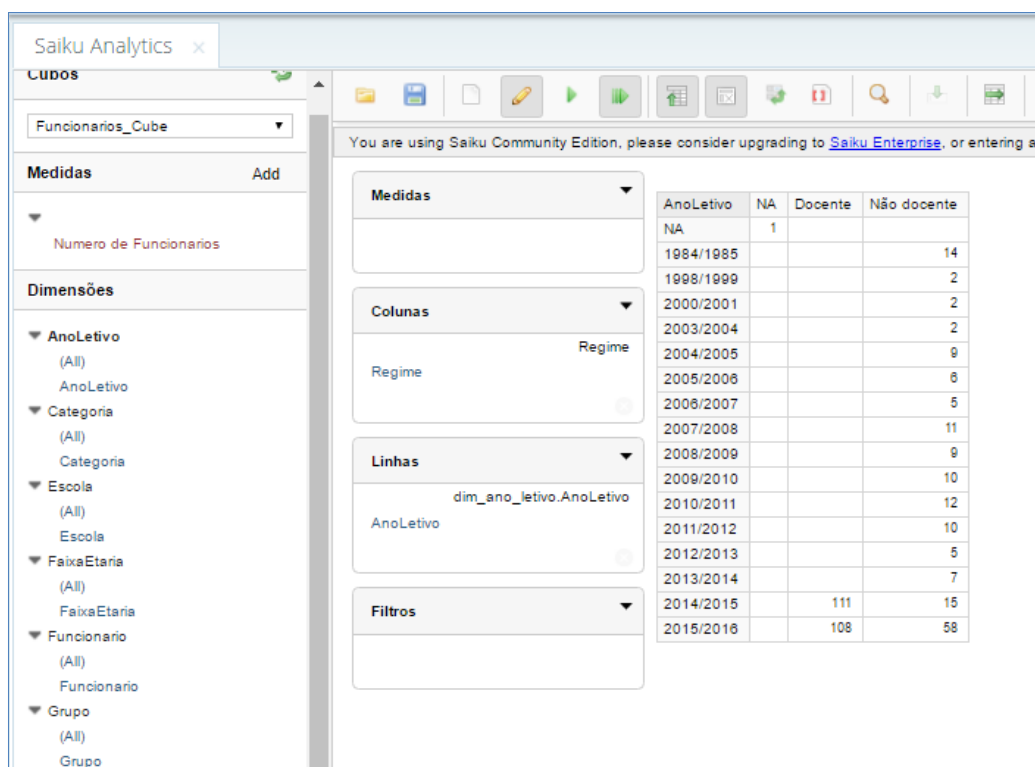


Figura 48: Ambiente de Trabalho Saiku – Matriz Cruzada

Na figura que se segue é possível visualizar a mesma exploração do cubo funcionários apresentada na figura acima, mas numa perspetiva gráfica.

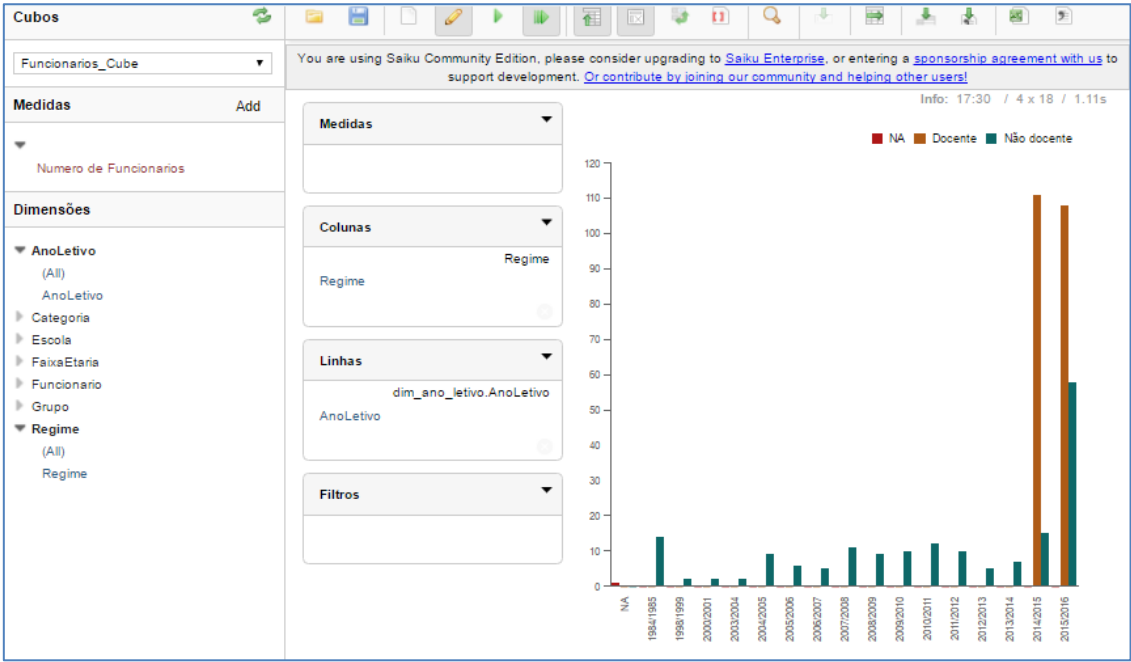


Figura 49: Ambiente de Trabalho Saiku – Exploração Gráfica

Uma das funcionalidades integradas no Saiku são as estatísticas básicas das pesquisas efetuadas ao cubo.

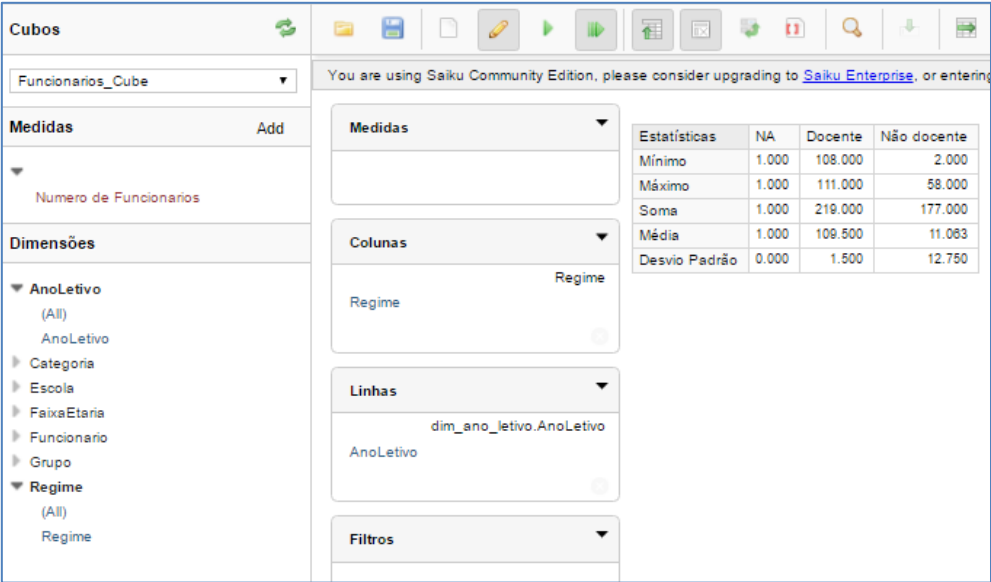
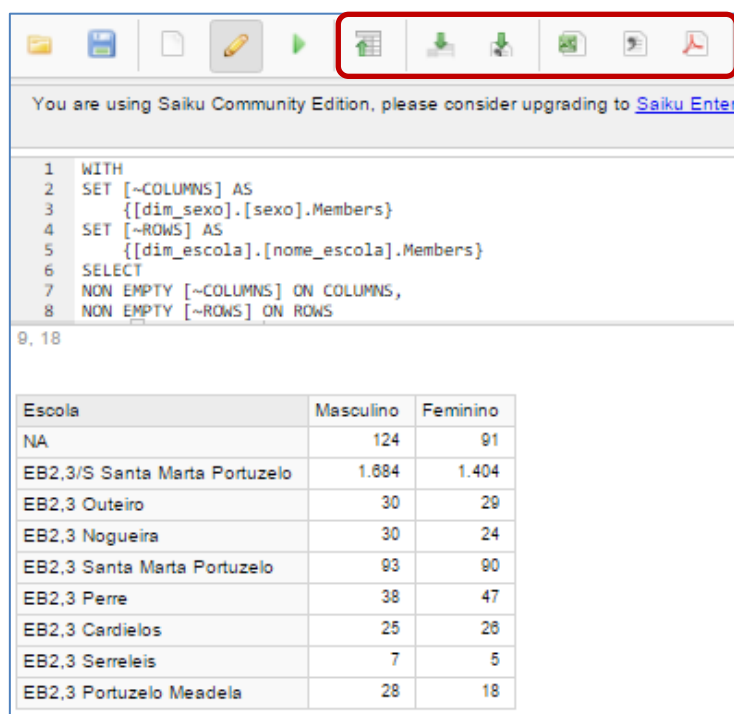


Figura 50: Ambiente de Trabalho Saiku – Estatísticas Básicas

Existe a possibilidade de exportar os gráficos para formatos como PNG, JPEG e PDF o que permite uma maior flexibilidade ao utilizador final.

Linguagem MDX: MDX é uma linguagem que foi desenvolvida para consultar dados armazenados nos cubos OLAP. Esta linguagem, idêntica à linguagem SQL, possibilita a realização de consultas multidimensionais, diferente da linguagem SQL que limita as consultas ao máximo de duas dimensões. As consultas podem ser exportadas em ficheiros do tipo csv, pdf ou xls. Para uma melhor compreensão, a imagem seguinte apresenta um exemplo prático.



You are using Saiku Community Edition, please consider upgrading to [Saiku Enterprise](#)

```

1 WITH
2 SET [~COLUMNS] AS
3   {[dim_sexo].[sexo].Members}
4 SET [~ROWS] AS
5   {[dim_escola].[nome_escola].Members}
6 SELECT
7   NON EMPTY [~COLUMNS] ON COLUMNS,
8   NON EMPTY [~ROWS] ON ROWS

```

9, 18

Escola	Masculino	Feminino
NA	124	91
EB2,3/S Santa Marta Portuzelo	1.684	1.404
EB2,3 Outeiro	30	29
EB2,3 Nogueira	30	24
EB2,3 Santa Marta Portuzelo	93	90
EB2,3 Perre	38	47
EB2,3 Cardielos	25	26
EB2,3 Serreleis	7	5
EB2,3 Portuzelo Meadela	28	18

Figura 51: Exemplo Query MDX e Respetivo Resultado

## 4.5 VISUALIZAÇÃO DE DADOS

A tarefa de *front-end* exige uma esquematização e um cuidado acrescido pois os *Dashboards* têm forte impacto num projeto de BI. Foi necessário ponderar as tecnologias de suporte a esta tarefa. Um dos pacotes de ferramentas mais conhecidos do Pentaho BI é a *suite* Community Tools (C\*Tools). O C\*Tools é um conjunto de ferramentas para construção de *Dashboards* para a *Suite* Pentaho BI *Open Source*, desenvolvido pela Webdetails.

Tendo como base a especificação de requisitos, foi desenvolvida uma aplicação web que permita à Comunidade Educativa extrair e visualizar informações de uma forma mais precisa e objetiva. A versão analisada no âmbito deste trabalho foi a *Community Edition*.

Esta plataforma apresenta o servidor de aplicações Apache Tomcat, a aplicação de administração do servidor, o ambiente de execução Java necessário para executar o Apache Tomcat e as respectivas aplicações web.

Para criação dos *Dashboards* foi utilizado o Pentaho BI Server 5.4.0 e fez-se recurso da *framework* CDF-DE que nos apresenta um interface web gráfico e intuitivo. Esta *framework* está a ser cada vez mais popular principalmente porque oferece facilidade na definição dos objetos e das variáveis, possibilita a atualização automática dos *Dashboards* conforme os dados, permite criar de forma simples *drop down list*, elementos de navegação, *templates* e possibilita também a criação de um ficheiro CSS.

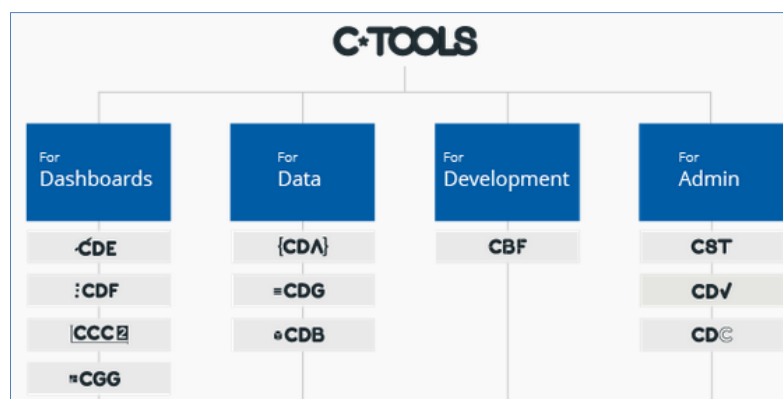


Figura 52: Componentes da suíte CTools

Fonte: It-novum

Com base na ilustração da figura anterior, a mesma é caracterizada por:

- **CDF** – *framework* que facilita a criação de *Dashboards* à base de componentes já existentes no ambiente de BI da Pentaho;
- **CDE** - ferramenta *user friendly* para construir, editar e pré-visualizar *Dashboards* no Pentaho. Tipicamente usa componentes CDF por trás. É basicamente um editor, sem ele teria de se codificar tudo à mão com CDF;
- **CDA** – componentes variados que dão acesso a diferentes tipos de **Datasources**.

No topo da aplicação CDE está presente um menu de opções (novo, gravar, atualizar). À direita estão presentes os menus técnicos, no ícon referente ao “*Layout*” é possível organizar o mesmo relativamente aos *Dashboards*, construindo instruções em HTML, ficheiros CSS e JavaScript.

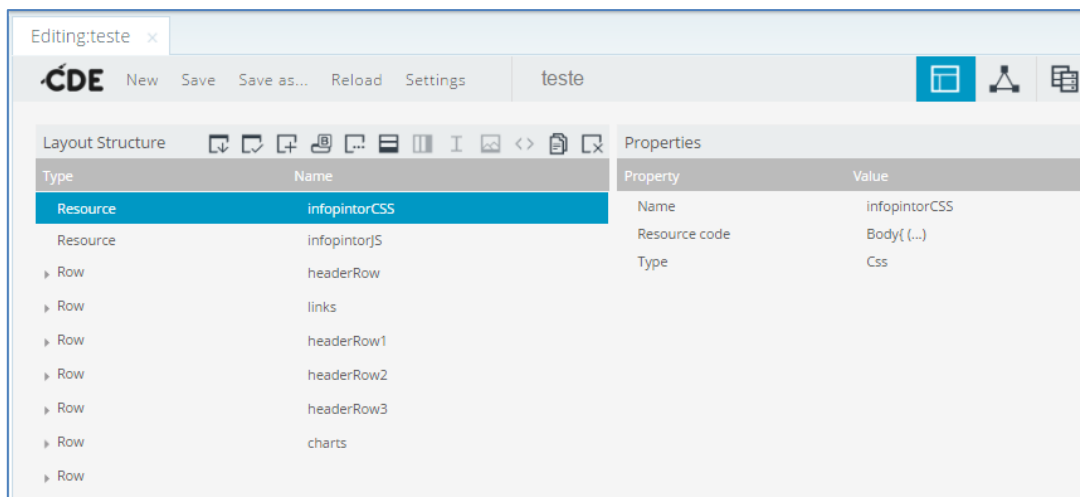


Figura 53: Vista do Menu *Layout*

No ícon “*Components*” são disponibilizados os diferentes tipos de gráficos a integrar na aplicação e é possível definir as caixas de seleção entre outros componentes.

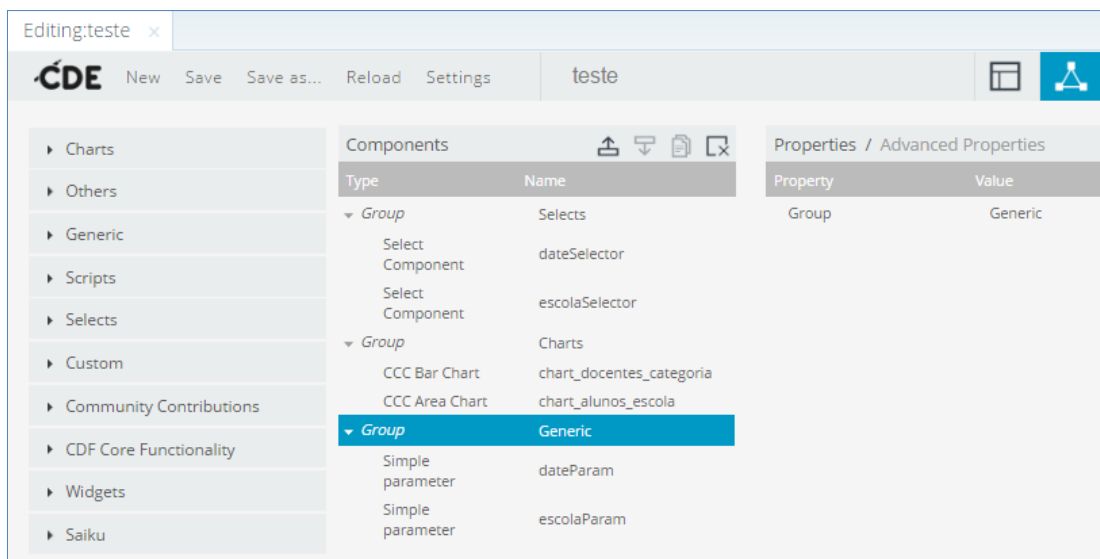


Figura 54: Vista do Menu *Components*

O ícon referente ao “*Data Sources*” apresenta elementos referentes às BD e às consultas. Nesta aplicação em concreto será feito recurso às consultas do tipo MDX.

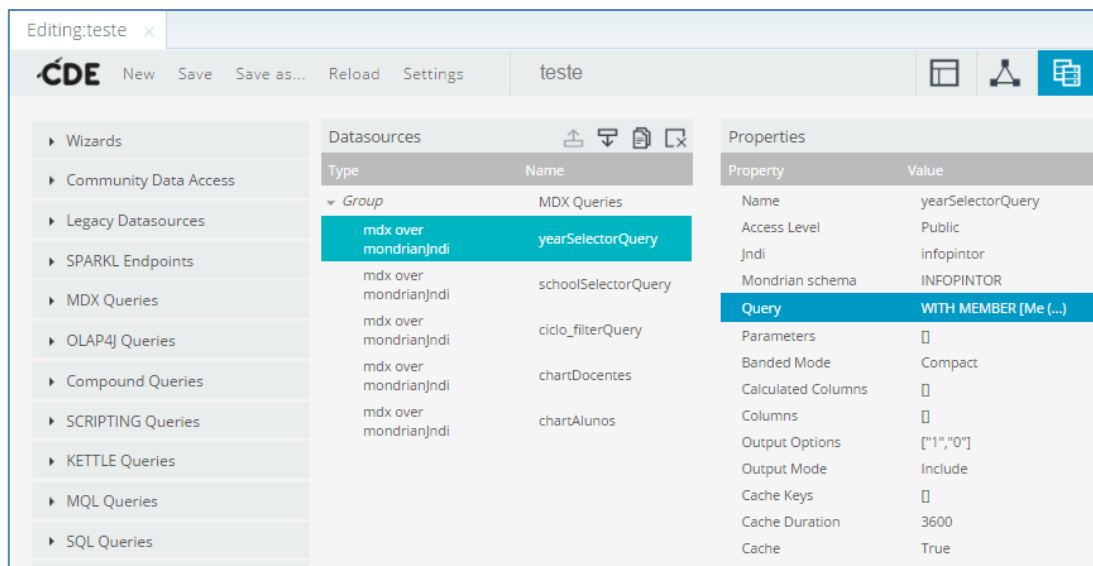


Figura 55: Vista do menu *Data Sources*

Como se pode averiguar na figura acima, no *Data Source*, existe a propriedade *query* onde são definidas as *queries* em MDX.

Os resultados apresentar graficamente decorrentes das *queries* podem ser muito complexos e variam muito com o pretendido pelo utilizador final. Não é possível pensar em todas as *queries* viáveis para um projeto, pelo que, estas vão sendo construídas dependendo dos filtros aplicar. Abaixo é visível um exemplo que retrata uma *query* que filtra os funcionários por tipo de regime (docente/não docente).

```
WITH
SET [~ROWS] AS
{[dim_regime_shared.dim_regime].[nome_regime].Members}
SELECT
NON EMPTY {[Measures].[numero_funcionarios]} ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [funcionarios]
WHERE {[dim_ano_letivo].[anoletivo].&[${dateParam}]*[dim_escola].[nome_escola].&[${escolaParam}]}
```

Figura 56: Exemplo de *Query* MDX

Por último o ícon “*Preview*” disponibiliza o resultado final do que foi construído, neste caso concreto a aplicação InfoPintor. Foram planeados cinco separadores determinantes, Escolas, Alunos, Professores, Funcionários e Explorar para dar resposta às várias visualizações gráficas solicitadas pelo Agrupamento.

Foram disponibilizados dois tipos de *Dashboards*, uns com uma estrutura fixa pré-definida e os *Ad-Hoc*, construídos a pedido com objetivos muito específicos, parametrizados pelo próprio utilizador. Estes últimos, presentes no menu “Explorar”, que

remete para a ferramenta Saiku. Com o acesso ao Saiku, a Comunidade Educativa poderá criar as consultas pretendidas para satisfazer as suas necessidades de informação.



Figura 57: Layout Final da Aplicação InfoPintor

## 4.6 MANUTENÇÃO E CRESCIMENTO

Na plataforma desenvolvida foram gerados os indicadores requeridos para a monitorização do sistema. No entanto, não invalida que futuramente sejam requeridos novos indicadores ou mesmo o aperfeiçoamento dos já existentes de forma a responder a novos objetivos que possam vir a ser estipulados. No âmbito desta dissertação, foram gerados todos os indicadores requeridos para os quais até ao momento já existiam informações registada. Isto é, tratando-se de um processo relativamente recente, nem todas as fontes de dados se encontram ainda disponíveis. Aliando os factos referidos anteriormente aos requisitos técnicos para a manutenção e crescimento da mesma, conclui-se que com os recursos disponíveis no Agrupamento é possível assegurar uma contínua disponibilidade, desempenho e expansão da plataforma de BI. No entanto, nesta fase inicial, pretendeu-se que a plataforma fosse utilizada pelos agentes de decisão, de forma a avaliar o seu desempenho.

## 4.7 APRECIÇÃO E AVALIAÇÃO GLOBAL DA PLATAFORMA

Com a realização deste projeto foi possível explorar a importância da integração da tecnologia de BI num Agrupamento de Escolas. Esta pode proporcionar benefícios para a instituição em questão, na medida que permite uma maior autonomia e flexibilidade dos



utilizadores relativamente à realização de análises de dados mais rápidas e mais precisas, apoiando a tomada de decisão.

A nível mais técnico e tendo em conta os requisitos impostos, foi utilizado o Pentaho CE como ferramenta de BI OS, pois foi a que melhor se enquadrou no âmbito deste projeto por ser dotada das seguintes características: construção e visualização de indicadores no *browser* e a integração de vários tipos de análise num só *Dashboard*. Assim, no mesmo *Dashboard* estão integrados diversos componentes o que para além de possibilitar ao utilizador visualizar e analisar a informação apresentada em vários contextos, permite também a realização de consultas *Ad-Hoc*. Não esquecendo ainda, o facto que o Pentaho CE permitir a construção de *Dashboards* com a característica de *bootstrap*, isto é, construir um *website* de modo que este se adapte ao dispositivo onde está a ser consultado: computadores tradicionais, *tablet* ou dispositivos móveis (*responsive web design*). Esta apresentação interativa perante o utilizador final e a possível adaptação do *layout* ao dispositivo no qual está a ser visualizada, permitirá resultar num aumento do número de utilizadores desta plataforma. Em termos globais, verifica-se que esta é bastante inovadora, uma vez que, são escassas as plataformas existentes que utilizam tecnologia de BI aplicada na área das Escolas/Agrupamentos e com estas características de acessibilidade e disponibilidade.

Em suma, neste capítulo foi descrita a solução implementada e foram especificados conceitos importantes para o desenvolvimento da aplicação proposta neste trabalho. Durante todo o processo de construção da aplicação, as reuniões com a Escola foram fundamentais para perceber as expectativas, ajustar conceitos e para compreender o que era pretendido ao nível da interação gráfica. Com esta aplicação novos cenários de tomada de decisão são mostrados o que proporciona ao órgão de gestão uma visão mais alargada e uma forma de atuação mais rápida e eficaz.

## 5. AVALIAÇÃO DOS RESULTADOS DO TRABALHO

### 5.1 INTRODUÇÃO

Neste capítulo serão apresentados e discutidos os principais resultados deste trabalho. A apresentação de todos os resultados seria extensa, pelo que são apresentadas conjugações viáveis e os respetivos resultados. Importa neste ponto retomar os objetivos e a questão de investigação deste trabalho, que passa por conceber uma plataforma integrada de gestão de informação *Open Source*, para um Agrupamento de Escolas.

Durante o desenvolvimento deste trabalho foram explorados conceitos e conhecimentos relacionados com BI, foram avaliadas ferramentas *Open Source* que respondiam às necessidades de implementação de um Sistema de BI e por último, foi implementado e avaliada a viabilidade deste tipo de solução num contexto Educativo. Nesta fase é feita a verificação se todos os requisitos foram cumpridos e se o projeto vai ao encontro do esperado pelo agrupamento.

Foram levados a cabo diversos testes: testes de interface de login e respetivos perfis, testes ao ETL, testes aos cubos e por fim testes à aplicação web. Foram efetuados testes à aplicação InfoPintor principalmente com os membros da Direção, Coordenadores dos Diretores de Turma, Equipa da Educação Especial, Secretaria e Psicóloga, sendo que as críticas foram muito positivas.

Foi também aplicado um inquérito de satisfação à usabilidade do sistema a todos os agentes de decisão do Agrupamento. O inquérito foi enviado via *email* a 14 professores que foram convidados a responder de forma confidencial. Os resultados globais dos questionários são mostrados no anexo 2.

### 5.2 TESTES

#### 5.2.1 INTERFACE LOGIN

A ferramenta BI *Server* é acedida via web. O controlo de acessos e respetivas políticas de segurança também estão presentes. Como medida de segurança, de maneira que o acesso aos dados estratégicos da Escola possa ser restringido, o Pentaho permite a gestão de utilizadores e de perfis que pode ser configurado na Área Administrativa.

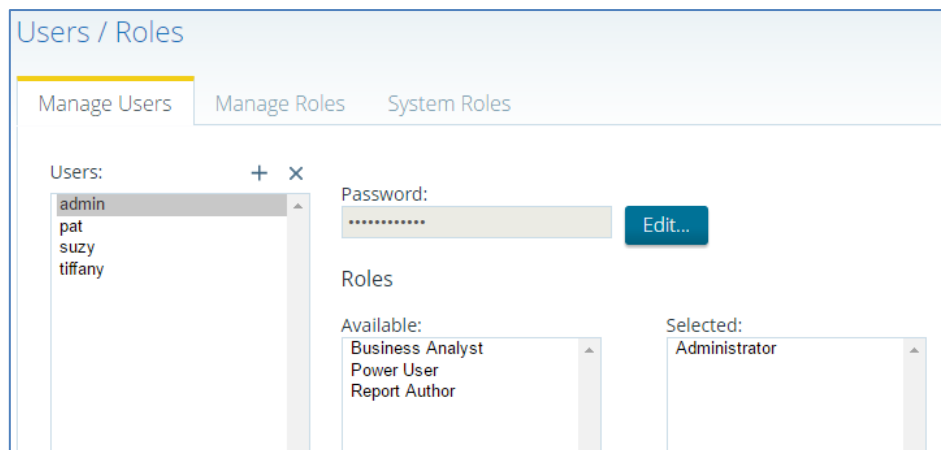


Figura 58: Gestão de Utilizadores

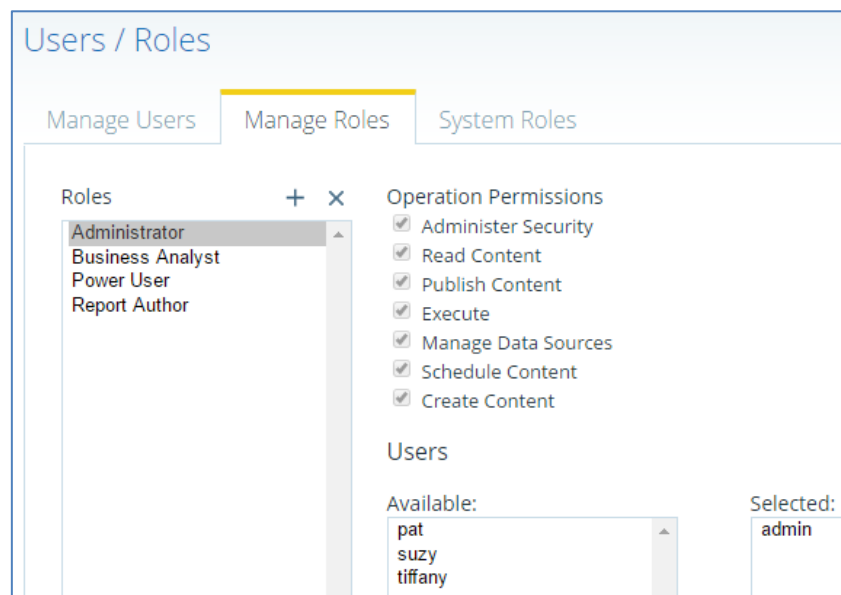


Figura 59: Gestão de Perfis de Utilizador

### 5.2.2 TESTES ETL/ DW

Na fase de testes referente ao ETL e DW foram comparados os resultados de consultas entre as BD operacionais e o DW com a finalidade de confirmar os dados carregados. Foram testados todos os componentes que integram o processo ETL à medida que estes iam sendo desenvolvidos. Com este tipo de testes foi assegurada não só a correta extração dos dados das bases de dados operacionais, mas também foi assegurada a consistência dos dados carregados no DW.

Como pode ser confrontado na figura abaixo representada, sempre que o processo ETL for compilado com sucesso surge um visto a cor verde. Na janela *Step Metrics* é possível

obter informações pormenorizadas dos dados lidos, escritos, atualizados, erros, tempo execução, entre outros.

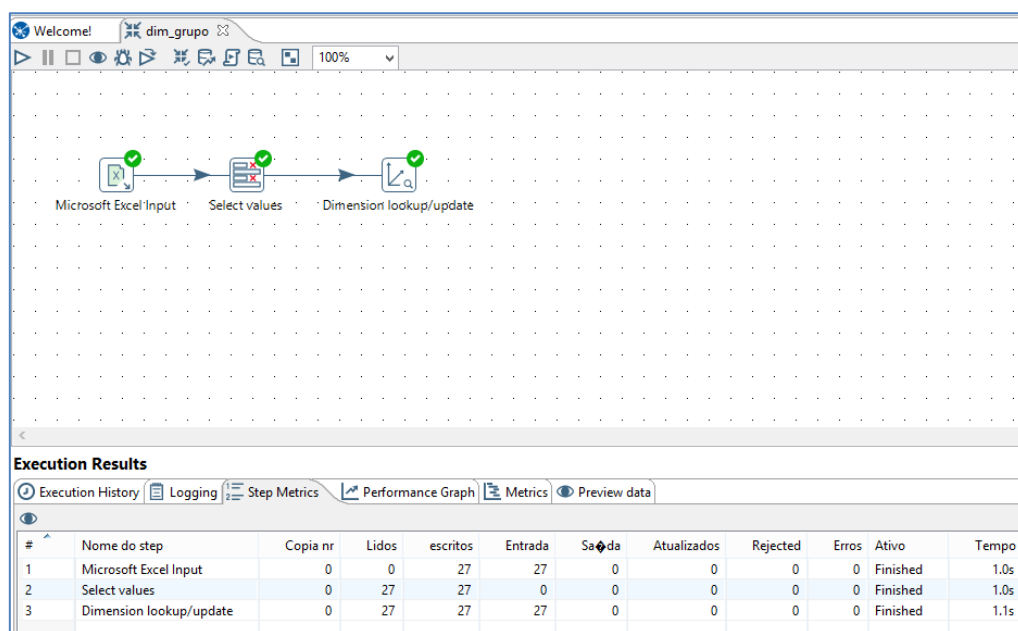


Figura 60: Testes ETL

Após esta verificação, foi executada uma *query* ao DW para confirmar o número de total de registos carregados. Foi também executada uma nova *query* às fontes de dados com a finalidade de conferir o número total de registos. Outros testes deste género foram levados a cabo, acrescentando novas condições, sempre com o objetivo de validar os dados tratados.

### 5.2.3 TESTES SAIKU ANALYTICS

Relativamente a este tipo de testes foi seguida a mesma linha de orientação do ponto anterior e foram feitos testes de comparação entre os dados do DW e dos cubos. Esta tarefa foi feita recorrendo à execução de *queries* no DW, e utilizando o explorador de cubos (Saiku) e *queries* MDX. As representações abaixo apresentadas mostram uma combinação de parâmetros elaborada no Saiku *Analytics*, onde é possível ver todos os funcionários por grupo de recrutamento.

Grupo	NA	Docente	Não docente
NA	1		170
100 - Educação pré-escolar		10	
110 - 1.º ciclo		29	
120 - Inglês - 1.º Ciclo		1	
200 - Português e Estudos		2	
210 - Português e Francês		3	
220 - Português e Inglês		3	
230 - Matemática e Ciências		5	
240 - Educação Visual		3	
250 - Educação Musical		2	
260 - Educação Física		3	
290 - Educação Moral		2	
300 - Português		11	
330 - Inglês		4	
350 - Espanhol		1	

Tabela 5: Validação dos Cubos no Saiku

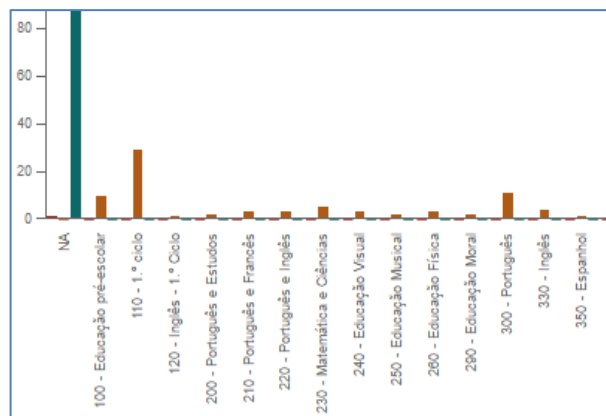


Figura 61: Validação Gráfica dos Cubos no Saiku

Run SQL query/queries on table infopintor.dim_funcionario:	
1	SELECT 'grupo', 'regime', COUNT('regime') FROM 'dim_funcionario'
2	GROUP BY 'grupo'

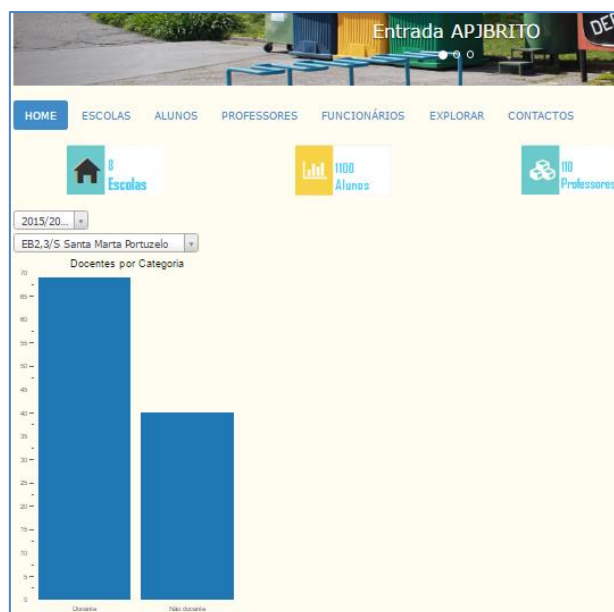
Figura 62: Validação através de Query SQL

grupo	regime	COUNT('grupo')
	NULL	0
0,0	Não Docente	170
100 - Educação pré-escolar	Docente	10
110 - 1.º ciclo	Docente	29
120 - Inglês - 1.º Ciclo	Docente	1
200 - Português e Estudos	Docente	2
210 - Português e Francês	Docente	3
220 - Português e Inglês	Docente	3
230 - Matemática e Ciências	Docente	5
240 - Educação Visual	Docente	3
250 - Educação Musical	Docente	2
260 - Educação Física	Docente	3
290 - Educação Moral	Docente	2
300 - Português	Docente	11
330 - Inglês	Docente	4
350 - Espanhol	Docente	1

Tabela 6: Resultado da Validação Query SQL

#### 5.2.4 TESTES APLICAÇÃO WEB

Os testes executados à aplicação InfoPintor incidiram principalmente na sua funcionalidade e fiabilidade da informação disponibilizada ao utilizador. Como exemplo podemos verificar na figura abaixo apresentada que o número de docentes para o ano letivo 2015/2016 para a Escola de Santa Marta é de 69 e não docentes de 40.



**Figura 63: Análise do Número Funcionários**

### 5.3 DISCUSSÃO DOS RESULTADOS

Os agentes de decisão passaram a ter informação privilegiada com a implementação do sistema de BI. A informação deixa de estar dispersa e passa a estar centralizada e acessível, aliado à vantagem do DW conter os dados históricos do Agrupamento. Efetivamente constatamos que a junção dos dados extraídos da aplicação Alunos e GPV numa única aplicação, utilizando dados históricos, permite uma visão alargada do Agrupamento através dos relatórios ou *Dashboards*.

O cubo OLAP permite conjugar variáveis que antes eram impensáveis, por exemplo se pretendermos obter de forma imediata todos os alunos subsidiados com escalão A na Escola Básica de Perre e Portuzelo, com o novo sistema a resposta é direta.

Algumas áreas foram consideradas prioritárias, principalmente, pelo seu grau de relevância dentro do Sistema de BI, com intuito de atender às dificuldades sentidas pelos agentes de decisão.

Até ao momento da implementação da nova aplicação, pessoal ligado à direção da Escola ou da secretaria, afirmam que quando precisam de fazer análise aos dados têm de os exportar para o Excel e, manualmente, fazer essa análise. Portanto, os recursos disponíveis até ao momento, apenas permitem saber, de forma generalizada certos dados, sendo que as características como *drill down* ou *roll up* são processadas de forma pouco eficiente em Excel.

As funcionalidades da nova aplicação têm como principal objetivo afastar gradualmente dos recursos usados até ao momento (relatórios em Excel) e, portanto, a procura de melhorias em termos de usabilidade e a garantia que todos trabalham com a mesma realidade.

Como vantagem implícita, é de salientar a possibilidade de gerar gráficos de uma forma simples e quase intuitiva foi o fator mais realçado, bem como os vários formatos em que é possível efetuar a exportação dos dados. O facto de a aplicação ser disponibilizada via Web, foi mais um fator enaltecido. A aplicação responde aos objetivos traçados facilitando a tomada de decisões é a opinião da Diretora do Agrupamento.

#### **5.4 DIFICULDADES ENCONTRADAS NA IMPLEMENTAÇÃO DO SISTEMA BI**

O primeiro contacto com a *suite* Pentaho foi uma das principais dificuldades encontradas neste projeto, bem como o estudo de todas as ferramentas usadas para o desenvolvimento e implementação do protótipo de DW. A esquematização do modelo em estrela também se mostrou complexo e difícil de concluir de modo a responder ao que era espectável para o Agrupamento.

#### **5.5 NOTAS FINAS**

Com a introdução de um sistema de BI no Agrupamento, foi possível minimizar a redundância de informações e possibilitou uma centralização dos dados permitindo assim responder em tempo útil a certas solicitações pedidas. Todas as opções tomadas na implementação mostraram-se capazes de solucionar o problema descrito.

O próximo capítulo apresentará as conclusões finais deste trabalho, as principais contribuições, assim como os trabalhos futuros.

## 6. CONCLUSÕES E TRABALHO FUTURO

### 6.1 CONCLUSÕES

O presente trabalho propôs um mecanismo de integração e consulta de informação para apoio à decisão num contexto escolar, neste caso concreto para o APJB. Numa primeira fase, na etapa de consciencialização observou-se que na Escola alvo do projeto existia a necessidade de obter indicadores de gestão. Neste âmbito verificavam-se que os registos são armazenados eletronicamente, no entanto não existe nenhum processo automático para extrair conhecimento útil e o disponibilizar em tempo real aos utilizadores finais (agentes de decisão).

Na etapa de sugestão foi proposto o desenvolvimento de uma plataforma de BI com vários módulos. Após várias sugestões e feita a análise das ferramentas de BI existentes no mercado, foi decidido, em consenso, com os agentes de decisão, que o *Pentaho Community Edition* seria a ferramenta mais adequada, assim como, também ficaram decididos e delineados os passos para o desenvolvimento do projeto. Existem algumas aplicações no mercado que permitem aos órgãos de gestão das Escolas analisar dados e tomar decisões céleres, no entanto ou são deficitárias ou têm preços pouco acessíveis. Assim, foi proposta uma solução baseada em ferramentas *open source* que respondesse às necessidades e objetivos do Agrupamento.

Na etapa de desenvolvimento foi então criado o artefacto requerido pela Escola. Neste sentido, a metodologia utilizada e exposta ao longo desta dissertação pode ser adotada por outras áreas da educação, outras instituições, ou mesmo por outras áreas transversais à área da educação.

Quanto à fase de avaliação a plataforma BI foi testada em ambiente real (utilizando dados reais) e posteriormente foram realizadas algumas reuniões com os profissionais da Escola envolvidos e utilizadores finais, de forma a avaliarem os resultados obtidos.

Com este estudo pretendeu-se também provar que efetivamente as soluções *open source* são viáveis e é possível implementar um projeto BI recorrendo unicamente a elas. A solução de BI para o APJBRITO apresenta inúmeras vantagens já referidas anteriormente no entanto, a Direção salientou a sua extrema importância para uma



possível autonomia da Escola e para a incorporação deste modelo no âmbito do Plano de melhoria.

Como outra conclusão deste trabalho, salienta-se que Integrar elementos de BI numa Escola não é tarefa simples, é necessário um conhecimento profundo do funcionamento do sistema Escola, é necessário um longo tempo de análise e compreender bem o funcionamento do Agrupamento. O APJB enriquece tecnologicamente dispondo de mais uma ferramenta educativa. A sua utilização por parte da Comunidade Educativa promove certamente o sucesso educativo.

## 6.2 TRABALHOS FUTUROS

O presente projeto apresenta uma ferramenta com muitas possibilidades de exploração e desenvolvimento futuro. Seguindo a mesma linha de trabalho poderia ser implementado em fases posteriores:

- Construção de um DM relativo a avaliações dos alunos;
- Verificar compatibilidades desta opção de BI com outros Agrupamentos;
- Promover a incorporação deste modelo no âmbito dos planos de melhoria do Agrupamento;
- Aplicação de outras técnicas de BI como o *Data Mining* assim como outras técnicas de visualização de dados;
- Otimização do *layout* dos *Dashboards*;

A disponibilização da informação via web, apresenta grandes desafios de segurança, colocando-se algumas questões éticas, sendo que a Escola não pode descurar a confidencialidade das informações que gere. Não foi esse o âmbito do estudo deste projeto, embora o acesso ao portal do Pentaho seja feito mediante login e *password*. Num trabalho futuro seria pertinente abordar o controlo dos acessos à informação, para garantir as políticas de privacidade, confidencialidade e segurança.

Seria também muito interessante e proveitoso averiguar o que mudou na gestão dos Agrupamentos que adotaram um sistema de BI como ferramenta de trabalho.

## REFERÊNCIAS

- Abreu, D. A. (2013). *Tecnologias da Informação Aplicada a Sistemas de Informação Empresariais*.
- Adamson, C. (2010). *Start Schema - The Complete Reference*. McGrawHill.
- Alves, W. P. (2009). *Banco de dados: teoria e desenvolvimento*. Érica.
- Araújo, E., Batista, M., & Magalhães, T. (s.d.). OLAP: Características, Arquitetura e Ferramentas. 2007.
- Bispo, A. F. (1998). Uma análise da nova geração de sistemas de apoio a decisão. *Dissertação de mestrado em Engenharia da Produção*.
- Caldeira, J. (2010). *Dashboards - Comunicar Eficazmente a Informação de Gestão*. Almedina.
- Chowdhury, G. G. (2014). Knowledge Organization or Information Organization? A Key Component of Knowledge Management Activities. *International Conference On Digital Libraries*.
- Chung, L., Nixon, B., Yu, E., & Mylopoulos, J. (2000). *Non-Functional Requirements in Software Engineering*. Springer US.
- Cockburn, A. (2005). *Escrevendo Casos de Uso Eficazes*. Porto Alegre.
- Codd, E. F. (1985). Is Your DBMS Really Relational?
- Costa, B. (1986). *Apoio à Tomada de Decisão segundo Critérios Múltiplos-Avaliação de Projectos e Decisão Pública*.
- Damiani, E. (2009). Open Source BI Adoption.
- Daradkeh, M. K. (2012). Information Visualisation to support Informed Decision-making under Uncertainty and risk. *Lincoln University Digital Thesis*.
- Davenport, T. H. (2002). *Ecologia da informação: porque só a tecnologia não basta para o sucesso na era da informação*.
- Diamantino, J. (2012). Sistema de acompanhamento do progresso do aluno.
- El-Sappagh, S. H., Hendawi, A. M., & Bastawissy, A. H. (2011). A proposed model for data. *Journal of King Saud University-Computer and Information Sciences*.
- Felber, E. W. (2005). Proposta de uma ferramenta olap num Data Mart comercial.
- Ferguson, R., & Lami, G. (2006). An Empirical Study on The Relationship Between Defective Requirements and Test Failures.
- Fernandes, J., Machado, R., & Seidman, S. (2009). A Requirements Engineering and Management Training Course for Software Development Professionals.
- Ferreira, M., Santos, J., & Serra, F. (2010). Ser empreendedor: Pensar, criar e moldar a nova empresa.
- Few, S. (2004). *Information Dashboard Design: The Effective Visual Communication of Data*.
- Geicer, J. (2003). *Mastering Data Warehouse: Relational and Dimensional techniques*.

- Golfarelli, M. (2009). *Open Source BI Platforms: a Funcional and Architectural Comparison*.
- Guerra, J., & Mendes, A. (2010). Sistema de Apoio à Decisão para a Gestão Escolar.
- Hall, D. A. (2011). *Applied Econometrics*. New York: Palgrave Macmillan.
- Inmon, W. h. (1996). *Building the Data Warehouse*.
- Järvinen, P. (1997). *Action Research is Similar to Design Science*.
- Jin, Z., & Fine, S. (1996). The effect of human behavior on the design of an information retrieval system interface. pp. 249-260.
- Katz, D., & Kahn, R. L. (1978). The social psychology of organizations. 2 ed. New York.
- Kay, R. (2010). *Data Cubes*.
- Kimball, R. (1998). *The Data Warehouse Lifecycle Toolkit*. Wiley.
- Koomey, J. (2011). Growth in data center electricity use 2005 to 2010. *The New York Times*.
- Lapa, J., Bernardino, J., & Figueiredo, A. (2014). A comparative analysis of open source business intelligence platforms.
- Machado, F. N. (2004). *Tecnologia e Projeto de Data Warehouse*.
- Maddocks, P., & Waring, T. (2005). Open Source Software implementation in the UK public sector: Evidence from the field and implications for the future.
- Manson, N. (2006). *Is operations research really research?*
- March, S., & Hevner, A. (2007). Integrated Decision Support Systems: A data warehousing perspective.
- Mike Uschold, R. J. (1999). A Framework for Understanding and Classifying.
- Pessoa, F. Á. (2002). Extensões de um cliente OLAP para consultas gerenciais a cubos numa BD multidimensional .
- Pires, R. F. (2011). O impacto da tecnologia da informação na geração de recursos competitivos nas empresas: uma abordagem baseada em estudos de casos.
- Ralph Kimball, M. R. (2002). *The Data Warehouse Toolkit*. Wiley Computer.
- Ranjan, J. (2009). Business Intelligence: Concepts, Components, Techniques And Benefits. *Journal of Theoretical*, pp. 60–67.
- Santos, A. P. (2013). *Institutos federais de educação: fontes de informação e gestão do conhecimento Federal institutes of education, science and technology of Brazil: information sources and knowledge management*.
- Santos, M. J. (2011). A Representação da Informação em Arquivos: Viabilidade de uso dos padrões utilizados na biblioteconomia.
- Shah, M. (2014). Impact of management information systems (MIS) on school administration: What the literature says. *Procedia - Social and Behavioral Sciences*.
- Shoshani, A. (1997). OLAP and statistical databases: similarities and differences.
- Speier, C., Vessey, I., & Valacich, J. S. (2003). The effects of Interruptions, Task Complexity, and Information Presentation on Computer-Supported Decision Making Performance.
- Studer, R. (1998). *Semantic Web Technologies*.

- Thorlund, G. H. (2010). *Business analytics for managers: Taking business intelligence beyond reporting*. Wiley.
- Vassiliadis, P. (2009). A Survey of Extract–Transform–Load Technology. *International Journal of Data Warehousing & Mining*.
- Vercellis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*. Wiley.
- Wang, H., & Hanna, S. (1997). Does Risk Tolerance Decrease With Age?
- Xu, L., Zeng, L., Shi, Z., He, Q., & Wang, M. (2007). Research on Business Intelligence in Enterprise Computing Environment. DOI: [IEEE10.1109/ICSMC.2007.4413870](https://doi.org/10.1109/ICSMC.2007.4413870).
- Yusuf, D. A. (1998). *Journal of Educational Review*.
- Zhu, B., & Chen, H. (2008). Information visualization for decision support. pp. 699-722.

# ANEXOS

## ANEXO 1 – MANUAL DE INSTALAÇÃO E CONFIGURAÇÃO DO SISTEMA

### Pré-Requisitos

- ✓ Sun Java Developer Kit (JDK)
- ✓ Ficheiro de instalação do Pentaho BI Server
- ✓ MySQL (opcional)

### Instalação

1. *Download* do JDK em: [oracle.com/technetwork/java/javase/downloads](http://oracle.com/technetwork/java/javase/downloads) e seguir os procedimentos de instalação.
2. Configurar a *path* para as variáveis de ambiente do java: adicionar duas linhas ao ficheiro com variáveis do sistema:

JAVA\_HOME : Diretório do Java: Neste caso > "C:\Program Files\Java\jdk1.7.0\_79"

JRE\_HOME : Diretório do JRE: Neste caso > "C:\Program Files\Java\jdk1.7.0\_79\JRE"

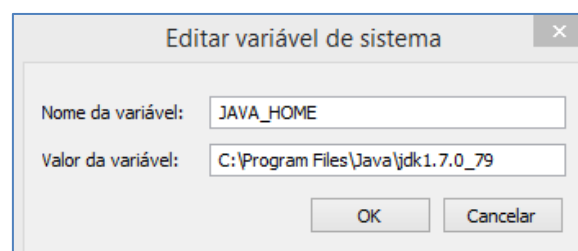
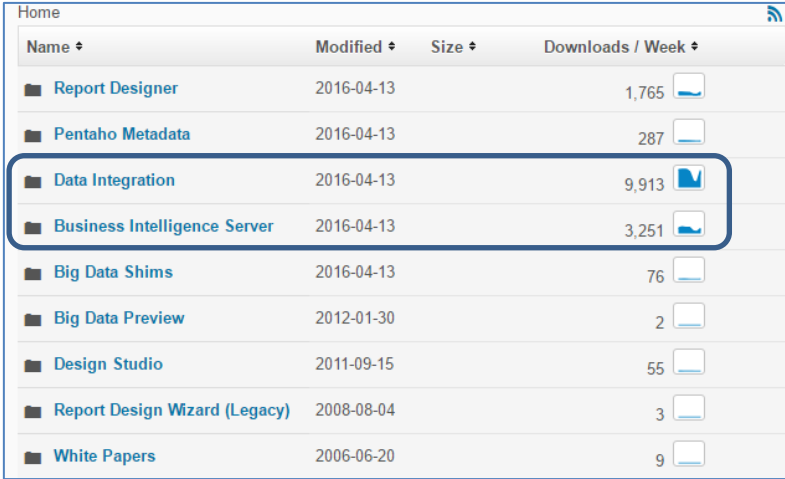


Figura 64: Variáveis de Sistema

3. Localizar a variável Path e inserir no final (;%JAVA\_HOME%/bin).
4. Instalação da Base de Dados MYSQL.
5. *Download* do Pentaho em: [sourceforge.net/projects/pentaho/files/](http://sourceforge.net/projects/pentaho/files/)
6. Efetuar *download* do BI Server e Data Integration.



Name	Modified	Size	Downloads / Week
Report Designer	2016-04-13		1,765
Pentaho Metadata	2016-04-13		287
Data Integration	2016-04-13		9,913
Business Intelligence Server	2016-04-13		3,251
Big Data Shims	2016-04-13		76
Big Data Preview	2012-01-30		2
Design Studio	2011-09-15		55
Report Design Wizard (Legacy)	2008-08-04		3
White Papers	2006-06-20		9

Figura 65: Download Pentaho

7. Extrair os ficheiros .zip para o diretório de instalação do Pentaho.
8. Arrancar o servidor Pentaho. Para iniciar o Pentaho é necessário executar o arquivo chamado start-pentaho.bat (Windows) ou start-pentaho.sh (Linux) na pasta "biserver-ce/". O Pentaho é executado sobre um Servidor Web, que por padrão é o tomcat. Deve aparecer a mensagem "INFO: Server startup in <tempo> ms" quando é finalizada a inicialização.

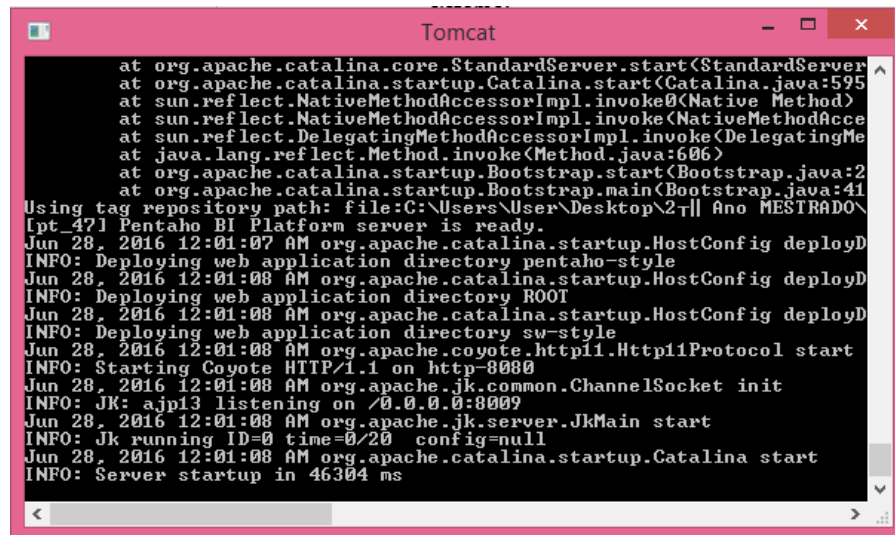


Figura 66: Consola Durante Execução Pentaho

9. Verificar se está operacional em: <http://localhost:8080/pentaho/> e efetuar login como Admin.

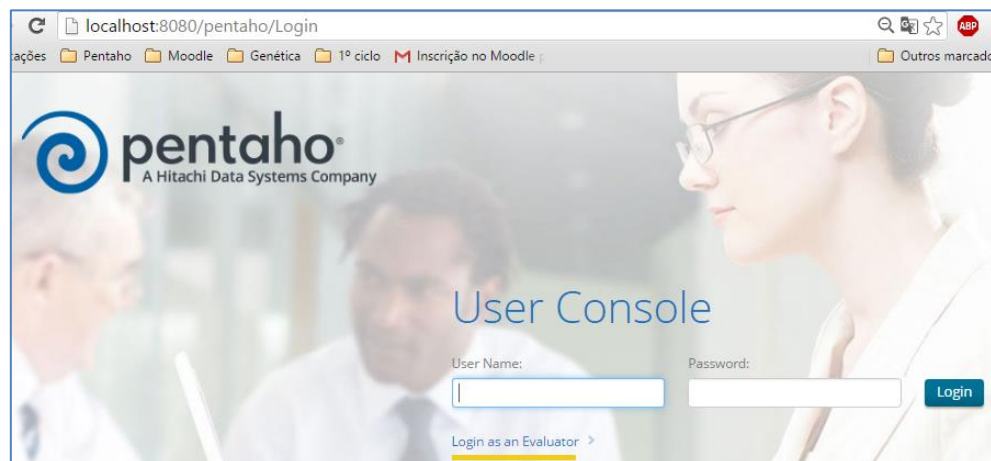


Figura 67: Janela de Login

10. Testar a ligação à BD em: Menu File --> Manage Data Source.

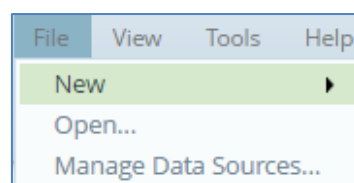
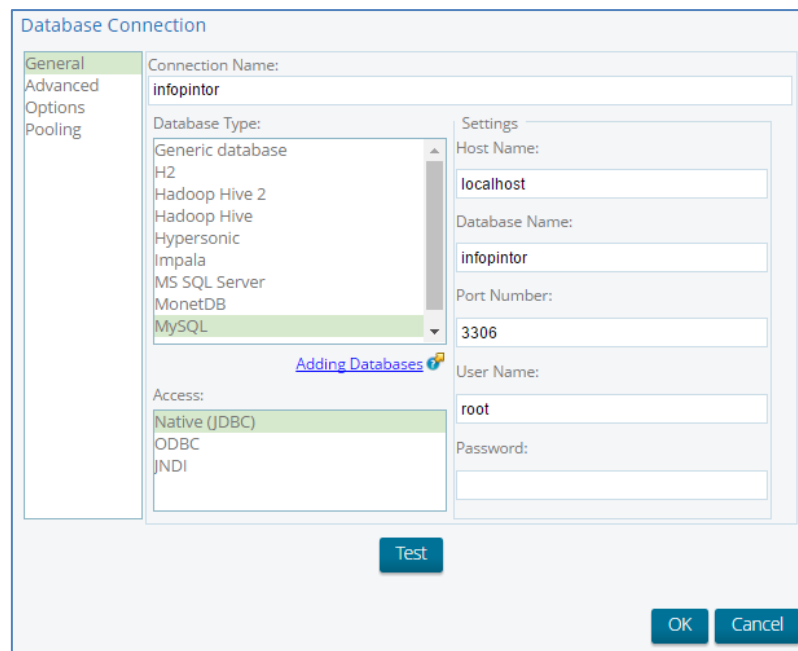


Figura 68: Menu File



**Figura 69: Ligação à BD**

11. Executar o ficheiro run\_chain.bat para correr o processo de ETL.
12. Aceder ao portal InfoPintor via web.

#### Outras Configurações:

1. Caso exista necessidade de executar o Pentaho Data Integration (PDI) – KETTLE:  
Executar o ficheiro em `\data-integration\spoon.bat`
2. Caso exista necessidade de executar o SAIKU: Instalação via Marketplace no menu Home --> Marketplace.
3. Verificar sempre os logs em: `biserver-ce/tomcat/logs/catalina.out`

## ANEXO 2 – INQUÉRITOS DE USABILIDADE DO SISTEMA

### Questionário

#### Questionário de satisfação do sistema InfoPintor

*\*Obrigatório*

**1. Grau de familiarização com as TIC? \***

*Marcar apenas uma oval.*

- ☐ Pouco
- ☐ Razoável
- ☐ Bom
- ☐ Excelente

**2. A aplicação InfoPintor tem uma apresentação gráfica agradável e legível? \***

*Marcar apenas uma oval.*

- ☐ Sim
- ☐ Não

**3. A navegação na aplicação é uma atividade... \***

*Marcar apenas uma oval.*

- ☐ Muito difícil
- ☐ Difícil
- ☐ Fácil
- ☐ Muito fácil

**4. De um modo geral, considera o acesso às informações da aplicação... \***

*Marcar tudo o que for aplicável.*

- ☐ Complicado
- ☐ Acessível
- ☐ Rápido
- ☐ Intuitivo

**5. Os dashboards apresentados na aplicação são: \***

*Marcar tudo o que for aplicável.*

- ☐ Claros
- ☐ Objetivos
- ☐ Confusos

**6. A aplicação InfoPintor respondeu a todas as questões pretendidas? \***

*Marcar apenas uma oval.*

- ☐ Sim
- ☐ Não
- ☐ Em parte



7. O acesso em dispositivos móveis é satisfatório? \*

Marcar apenas uma oval.

- ☐ Sim  
☐ Não

8. Penso que precisaria do apoio técnico para conseguir usar a aplicação na perfeição \*

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

9. A aplicação dá prioridade à informação importante? \*

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

10. As ligações externas (ex: Saiku) são úteis? \*

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo fortemente

11. Atribui uma classificação de qualidade global a esta aplicação, atendendo a todos os parâmetros analisados de...

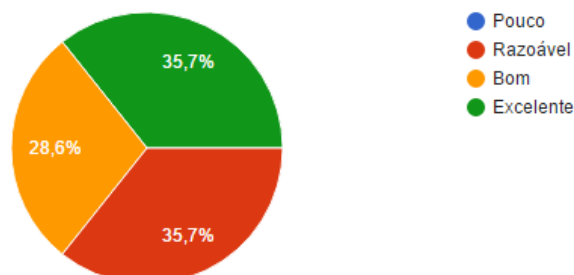
Marcar apenas uma oval.

- ☐ Excelente  
☐ Muito bom  
☐ Bom  
☐ Razoável  
☐ Fraco

Com tecnologia  
 Google Forms

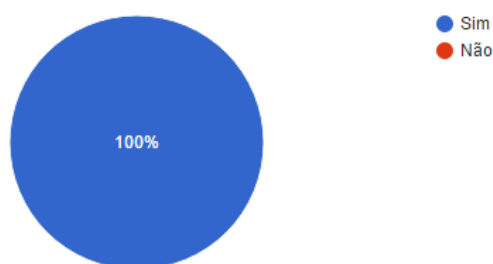
## Resultados Apurados

Grau de familiarização com as TIC? (14 respostas)

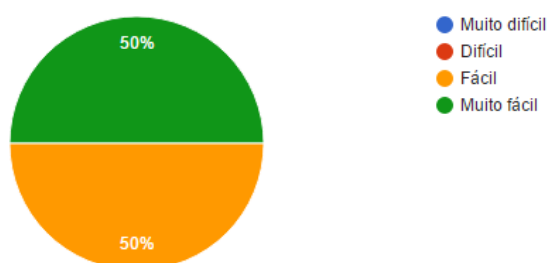


A aplicação InfoPintor tem uma apresentação gráfica agradável e legível?

(14 respostas)

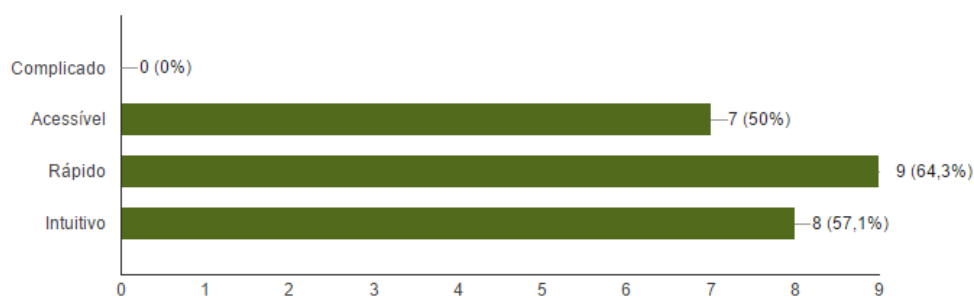


A navegação na aplicação é uma atividade... (14 respostas)

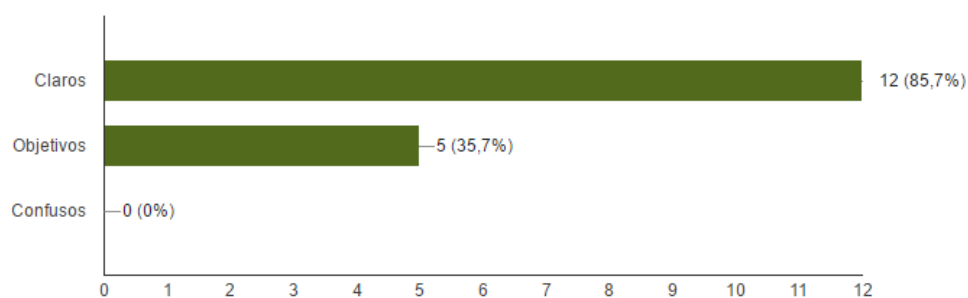


De um modo geral, considera o acesso às informações da aplicação...

(14 respostas)

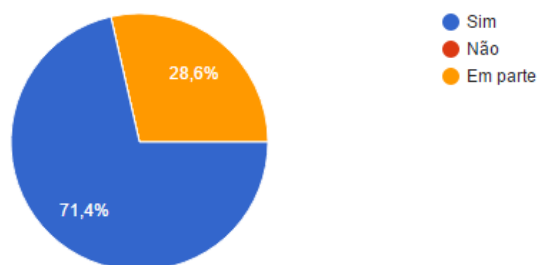


Os dashboards apresentados na aplicação são: (14 respostas)

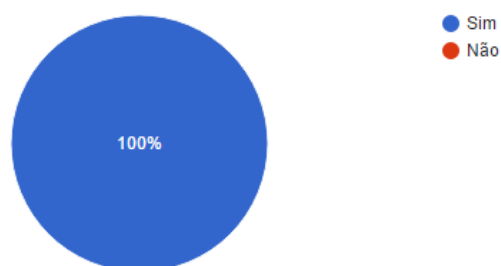


A aplicação InfoPintor respondeu a todas as questões pretendidas?

(14 respostas)

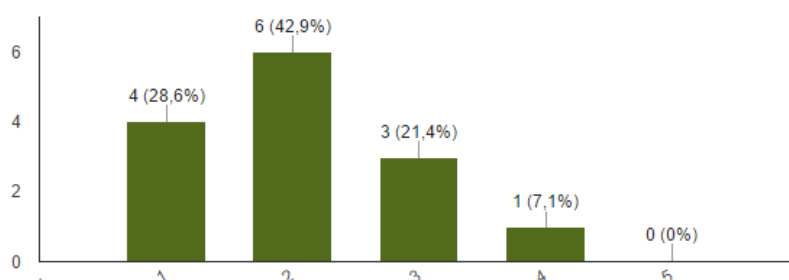


O acesso em dispositivos móveis é satisfatório? (14 respostas)

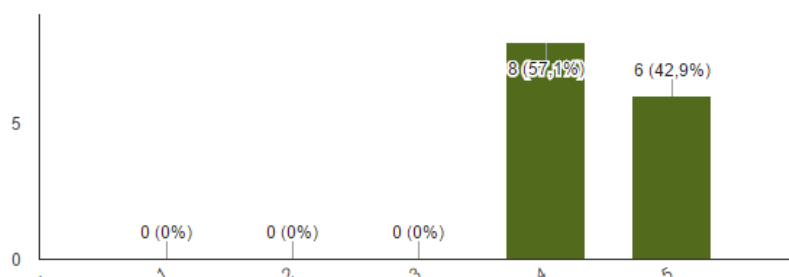


Penso que precisaria do apoio técnico para conseguir usar a aplicação na perfeição

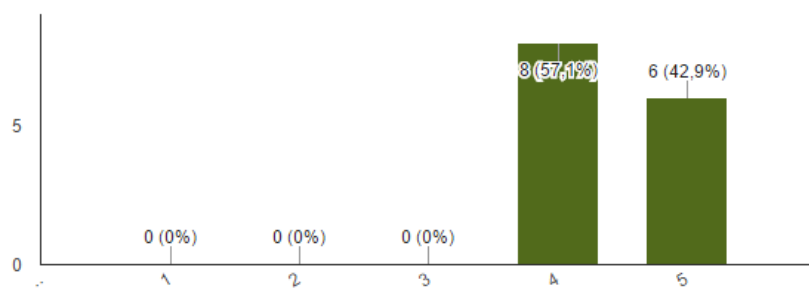
(14 respostas)



A aplicação dá prioridade à informação importante? (14 respostas)



As ligações externas (ex: Saiku) são úteis? (14 respostas)



Atribui uma classificação de qualidade global a esta aplicação, atendendo a todos os parâmetros analisados de...

(14 respostas)

